Title
Lessons in Plant Cell Wall Degradation by the Ascomycete Model Fungus, Neurospora crassa

Author
Wu, Vincent

Publication Date
2017

Peer reviewed|Thesis/dissertation

Lessons in Plant Cell Wall Degradation by the Ascomycete Model Fungus, *Neurospora crassa*

by

Vincent Wei-Xiang Wu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor N. Louise Glass, Chair
Professor Rachel Brem
Professor Henrik Scheller
Professor Nicholas Ingolia

Fall 2017

Abstract

Lessons in Plant Cell Wall Degradation by the Ascomycete Model Fungus,

Neurospora crassa

by

Vincent Wei-Xiang Wu

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor N. Louise Glass, Chair

In order to mitigate the negative environmental effects of anthropogenic global climate change, renewable resources must be developed to replace fossil fuels. One potential fossil fuel replacement platform utilizes plant biomass from agricultural waste as a feedstock to generate fuels and high value products. Plant cell wall polysaccharides including cellulose, hemicellulose and pectin are broken down into fermentable sugars that can be converted to compounds of interest by microorganisms. Filamentous fungi are capable of both depolymerizing polysaccharides and converting fermentable sugars to high value products. However, further engineering is required to generate strains of filamentous fungi that can efficiently secrete both polysaccharide-degrading enzymes and/or synthesized compounds of interest. This engineering is only possible through a better understanding of the metabolic biology of these organisms. This dissertation describes the mechanisms by which the model filamentous fungus *Neurospora crassa* regulates and produces the enzymes required for depolymerization of plant cell wall polysaccharides.

We investigated pyroglutamate, a unique post-translational modification, and have determined the mechanism of catalysis, as well as its role in the secretion of plant cell wall degrading enzymes. Pyroglutamate modification is catalyzed by two glutaminyl cyclases localized in the endoplasmic reticulum. Secreted proteins with either an N-terminal glutamine or glutamate are likely to have this terminal amino acid converted to a pyroglutamate. Pyroglutamate provides proteins with protection from native *N. crassa* proteases. The modification also provides structural stability for enzymes within glycosyl hydrolase family 7, including CBH-1 and other important cellulases. The amino acid sequences of glutaminyl cyclases are highly conserved within the fungal kingdom, including critical amino acids required

for catalysis and substrate binding. These data indicate that our findings regarding the role of pyroglutamate in *N. crassa* may be broadly applicable to other fungal genera.

Transcriptional profiling of *N. crassa* exposed to various components of the plant cell wall resulted in the identification of a number of candidate transcription factors and sugar transporters involved plant cell wall deconstruction. Results from these experiments also suggest that *N.* crassa is responsive to a diverse set of signaling sugars. These sugars induce transcription of plant polysaccharide lytic enzymes. The responses to these sugars can explain transcriptional responses of *N. crassa* to parent plant cell wall polymers from which sugars are derived.

Analysis of candidate transcription factors revealed a complete set of regulators responsible for positively regulating expression of plant cell wall degrading enzymes in the *N. crassa* genome. Regulation of hemicellulytic and pectinolytic enzymes is governed by PDR-1, PDR-2, and ARA-1 along with the previously described xylanase regulator XLR-1. These transcription factors are conserved homologs of *RhaR, GaaR* and *Ara-1,* respectively, from *Aspergillus* and other species. We found additional transcription factors that we named HEM-1 and MAN-1 that play a minor role in the positive regulation of hemicellulases. Transcriptional data from engineered strains lacking these transcription factors, along with data from Δ*clr-1, Δclr-2* and Δ*xlr-1,* indicate that these sugar responsive regulators have overlapping regulons. Surprisingly, none of these transcription factors affect the expression of one another.

DNA affinity purification sequencing (DAPseq) was used to assay direct DNA-binding of a number of *N. crassa* transcription factors. This assay is comparable to chromatin immune-precipitation sequencing (ChIPseq). Direct binding data of XLR-1 and CLR-2 from both DAPseq and ChIPseq indicate that the same core set of hemicellulases and cellulases are directly regulated by the two transcription factors, respectively. Additionally, DAPseq binding sites for nitrogen and sulfur metabolism regulators NIT-2, CYS-3 and NUC-1 confirm previously described targets for these transcription factors.

Direct binding data from DAPseq provides new insights into the mechanisms of function of the carbon catabolite repressor CRE-1. Contrary to previous research, DAPseq data shows that CRE-1 represses the expression of cellulases through the repression of cellobiose transporters *cdt-1, cdt-2,* and *cbt-1.* Repression of hemicellulases and pectinases expression may occur in a similar manner.

DAPseq data also revealed the mechanism of VIB-1 influence in plant cell wall degradation. VIB-1 not only directly regulates expression of genes involved in vegetative incompatibility, phosphatase production and protease production described in previous literature, but VIB-1 also directly binds the promoters of *clr-2, xlr-2, pdr-2, ara-1* and others key regulators of polysaccharide deconstruction. Direct binding of VIB-1 to the *clr-2* promoter is consistent with previously published data. VIB-1 directly regulates many genes involved in nutrient acquisition of all types and may be a key regulator that integrates signals from many types of nutrient limitation.

The complementary CRE-1 and VIB-1 data provide a new model for induction of plant cell wall degrading enzymes in ascomycete fungi. This model suggests that

VIB-1 may be responsible for sensing of extracellular plant cell wall polysaccharides by promoting basal transcription of polysaccharide degrading enzymes capable of releasing signaling sugars that increase production of select enzymes. The regulation of transporters by CRE-1 occurs upstream of this VIB-1 induction, allowing the two transcription factors to work cooperatively.

Table of Contents

List of Figures

Acknowledgements

Louise Glass has made this body of work possible with her consistent support and guidance. Louise has always pushed me to be a better scientist from conceptual ideas to basic experimental set ups. She has connected me with the people and techniques that were required to accomplish the work described in this

dissertation. Most importantly she has created a thriving lab environment where all members can learn from, depend on and enjoy the company of one another. It has been a great joy to be a part of this atmosphere.

A number of lab members have been incredibly important to my development as a scientist. They have helped me directly with techniques and lessons that have made this body of work possible. They have taught me all I know about molecular biology, fungal biology, genetics and computational biology. These individuals include David Kowbel, Jason Liu, Lori Huberman, Lina Qin, James Craig, Darae Jun, Sam Coradetti, Yi Xiong, Christy Roche, Morgann Reilly, Monika Fisher and Trevor Starr. Other lab mates who have contributed much advice and discussion include Jens Heller, Gabe Rosenfield, Asen Daskalov, Pedro Goncalves, Timo Schuerg and Jiuhai Zhao, Wilfried Jonkers, Javier Palma-Guerrero. David, James, Jason, Lori, Sam and Lina have been especially important mentors and role models during my time here at Berkeley.

Craig Dana and Christy Roche were critical for the work on pyroglutamate modification. Their work initiated the project, and discussion and planning with them allowed the project to continue and succeed. Craig, whom I collaborated directly, was wonderful source of ideas and discussion.

The research presented regarding pectin metabolism would not have been possible with our collaborators at the Technical University of Munich, Philipp Benz and Nils Thieme. I am grateful to have been able to take part in their exploration of *pdr-1* and pectin metabolism. Nils and Philipp have been a joy to work with and discussions with them have sparked the formation of many new ideas and leads.

Lastly, the Joint Genome Institute along with the folks in the fungal group and genomic technologies group including Igor Grigoriev, Sara Calhoun, Asaf Salamov, Vasanth Singan, Ronan O'Malley, Juna Lee and Remo Monti have all been critical to the majority of work presented in this dissertation. I have greatly benefitted from all the help and work on the transcriptome profiling project and much more. Sara has been especially helpful in using her expertise to analyze explore much of the RNAseq data that we obtained. The genomics technology group including Ronan O'Malley, Juna Lee, and Remo Monti were all instrumental in completing all of the DAPseq experiments. I am very grateful to Ronan for allowing me to use his resources to adapt DAPseq to fungal transcription factors. Juna Lee was exceptionally gracious in teaching me and providing all that was necessary to complete these experiments.

Chapter 1. **Filamentous fungi as a platform for high value product production from agricultural waste**

1.1. Introduction

Prevention of increased global anthropogenic climate change requires an immediate switch to renewable resources from current reliance of fossil fuels. Atmospheric carbon dioxide concentration has increased over 100 ppm in the last century to annual mean value of 400 ppm, an alarming number when compared to the natural fluctuation between 180 and 280 ppm over the past several million years (1). This increase in carbon dioxide is primarily caused by the combustion of fossil fuels (2). The increase in global CO_2 levels is currently causing an increase in global temperatures with catastrophic consequences to communities across the globe (3). To combat further production of atmospheric CO_2, carbon negative and neutral technologies must be developed and utilized in place of fossil fuels. One such technology employs microorganisms to produce fuels and high value products from carbon neutral feedstocks. This technology harnesses microbial decomposition of organic material to convert agricultural waste to useful compounds and products in a carbon neutral manner.

Fungi are becoming increasingly important in this field of industrial biotechnology. Various fungi and bacteria are already widely used to produce chemicals, polymers, enzymes and fuels, but filamentous fungi have been increasingly exploited to generate such products. Filamentous fungi possess several important attributes that confer advantages over yeasts and bacteria. These fungi form extended networks of cells capable of rapid expansion into diverse substrates, and their unique physiology also boasts a robustness and secretory capacity unmatched by other microorganisms (4, 5).

Historically, filamentous fungi were used for the production of enzymes and metabolites. Products include antibiotics, organic acids, pigments and plant biomass degrading enzymes (6). Plant biomass degrading enzymes are important for food and textile industries to modify or deconstruct polymers such as starch and cellulose. To increase production efficiency, research has focused on increasing protein yields in these organisms. Mutagenesis and selection has been the dominant approach. Hyper secretion phenotypes were repeatedly selected for over years of mutagenesis and selection. This process has resulted in the hyper-secreting strains used in various industries. Commonly utilized species include *Trichoderma reesei*, *Aspergillus niger* and *Myceliopthroa thermophila*, but additional taxa are being developed for industrial use (4).

Engineering strains that can directly consume plant feedstocks and produce products of interest is of great interest to biofuels researchers. Ideally, feedstocks would consist of low cost agricultural waste that can be directly consumed by fungi, while those fungi simultaneously produce and secrete a product of interest (7). Alternatively, plant cell wall degrading enzymes (PCWDE) can be harvested from fungi and utilized to depolymerize plant material along with chemical pretreatment. Freed sugars can be fermented by a secondary organism to produce final products of interest, though this is less efficient (8). Strain engineering for increased protein

secretion is required for both strategies. While the advent of molecular biological tools has made strain engineering more efficient, the underlying biology of protein production and secretion is still poorly understood.

In this dissertation I describe efforts to gain a better fundamental understanding of how filamentous fungi sense and respond to different types of plant matter, with a focus on expression and secretion of PCWDEs. This information may allow for more rational engineering of these organisms. The research presented takes place in the model filamentous fungus *Neurospora crassa.* This organism was chosen due to its ability to consume the plant cell wall, as well as for its ease of study. An extensive set of molecular biological tools has been developed after almost a century of research using *N. crassa* (9). The work described in this dissertation focus on a post-translational modification to PCWDEs, as well as the transcriptional regulation behind PCWDE production in *N. crassa*.

1.2. *Neurospora crassa* as a model organism

Neurospora is a genus of ascomycete fungus within the class Sordariomycetes. It has been studied since the mid 1800's as a contaminant of French bakeries, known early as the "red bread mold." A visually similar fungus was known to be a the biological agent in the fermentation of oncom, fermented soybean or peanut cakes a traditional staple food of West Javan cuisine. Early on, these orange spored fungi were know as *Monilia sitophila*. Work conducted by Shear and Dodge at the U.S. Department of agriculture separated the *Monilia sitophila* group into three species characterized by their divergent sexual structures. These species were *Neurospora crassa, Neurospora sitophila* and *Neurospora tetrasperma* (10). A picture of *N. crassa* is shown below.

Neurospora gained fame when geneticists Beadle and Tatum saw the advantages of *Neurospora* as a genetic system in comparison to the fruit fly *Drosophila*. They published the first paper in 1941 on the isolation of three biochemical mutants of *Neurospora,* two requiring vitamin B_6 and third mutant requiring p-amino benzoic acid (11). These early mutants were generated by X-ray treated cells plated onto selective enrichment media .

Advances in the development of *Neurospora* as a system for genetic studies were furthered by Stanford researcher David Perkins in the early 1950's.

Neurospora continued to provide insights into nutrient metabolism fundamentally different from information gained from bacteria. However, this work was continued in *Saccharomyces cerevisiae* during 1960's and 1970's when the yeast proved to be a more tractable organism. The use of *Neurospora* as a genetic system experienced a revival with the advent of molecular biological techniques. In the past three decades it has become a leading organism in the study of photobiology and circadian rhythms, a model for plant pathogenic fungi, an opportunity to explore evolution of heterothallism and homothalism and a model for exploring plant cell wall degradation by filamentous fungi (10).

1.3. Plant cell wall polysaccharides

The plant cell wall is a composed of a complex and integrated set of polysaccharides that can vary greatly across tissue type and across plant species (12). Plant cell wall polymers are divided into 4 classes: cellulose, hemicellulose, pectin and lignin (13). Cellulose, hemicellulose and pectin are polysaccharides, while lignin is a class of organic polymers typical of woody biomass. As *N. crassa* is unable to utilize lignin as a carbon source, this work focuses on cellulose, hemicellulose and pectin. These three polymers can further divided into the specific polymers described below.

Cellulose

Cellulose is the most recalcitrant and most abundant cell wall polysaccharide. It is composed of β, 1-4 linked D-glucose residues arranged in linear chains. Single chains align in parallel, connected by hydrogen bonds, to form microfibrils with extreme tensile strength. These microfibrils are the backbone of the plant cell wall, providing structure and rigidity.

Hemicellulose

Hemicelluloses are a class of polysaccharides that are composed of a number of different polymers. Hemicelluloses are generally abundant in the sugar xylose, and serve as secondary polymers that hold cellulose microfibrils together in the plant cell wall (13). Hemicellulose represents about 20-35% of lignocellulosic biomass (14). Several common hemicelluloses are described below.

Xylan

Xylan is a common hemicellulose composed of a backbone of β 1-4 linked xylose and varying side chains that contain a wide range of sugars including, but not limited to, arabinose, galactose, glucuronic acid and acetate. Certain xylans, like arabinoxylan, glucuronoxylan, and glucurononarabinoxylan, have heteropolymeric backbones. The ratio of xylose to other sugars can be highly variable, depending on the plant species. For instance, birchwood xylan contains ~89% xylose, while rice bran xylan contains 46% xylose and 45% arabinose (14).

Xyloglucan

Xyloglucan is a ubiquitous plant cell wall polymer with a β 1-4 linked D-glucose backbone and xylose side chains. It is the most abundant hemicellulose in the dicot primary cell wall, and makes up 20-25% of the dry weight (15). Due to its ability to bind to cellulose via the hydrogen bonding of its back bone, xyloglucan has been historically thought of as a crucial hemicellulose (16).

Mannan

Mannan is polymer with a backbone of β 1-4 linked D-mannose residues. Mannans have been well characterized in their role as seed storage compounds, but are also found in variable amounts in all plant cell walls. Mannans are abundant in early diverging plants such as mosses and lycophytes, but have been replaced by other hemicelluloses in spermatophytes (15).

Pectin

Pectin is a class of polysaccharide characterized by the abundance of D-galacturonic acid (GalA), and is of heterogeneous structure. Pectins are integral to plant cell growth and morphogenesis. Pectins are crosslinked with hemicellulose and cellulose, and affect cell wall flexibility and strength. The two most common forms of pectin are homogalacturonan (HG) and rhamnogalacturonan I (RG1), but other forms of pectin exist. HG and RG1 are described below.

Homogalacturonan (HG)

HG is a pectin with α 1-4 linked GalA, and can account for greater than 60% of the pectins in the plant cell wall. GalA in HG can be methyl-esterified, but methyl-esterification is variable from source to source. Un-methyl-esterified GalA can interact with calcium ions and induce gel formation common in fruit, and is also observed in plant cell walls (17). Side chains of xylose and acetate are common.

Rhamnogalacturonan I (RG1)

RG1 is more complex than HG and is composed of alternate GalA and rhamnose residues. GalA is bound to rhamnose by α-1,2 and α-1,4 linkages. RG1 has a diversity of side chains including arabinan and galactan or arabinogalactan, that branch from rhamnose backbone residues. Arabinose residues are generally linked by α-1,5 linkages, while galactose residues are linked by β-1,4, β-1,6 linkages or β-1,3 linkages (17).

1.4. Fungal enzymes for plant cell wall degradation

Fungi employ a wide range of enzymes to depolymerize the plant cell wall. The types and numbers of these enzymes can vary greatly between fungi, depending on the ecological niche and preference of each species. In general, plant associated filamentous fungi are able to degrade and metabolize a diversity of polysaccharides and sugars, due to the complexity of the plant cell wall. This section summarizes the enzymes present in the *N. crassa* genome that will be further referenced in the upcoming chapters of this dissertation. Enzymes are organized by the classes of

polysaccharides they can degrade, and are manually curated from the set of the carbohydrate active enzymes (CAZys) present in the *N. crassa* genome.

Cellulose degrading enzymes:

N. crassa has an extensive array of cellulose degrading enzymes. These enzymes make up the largest portion by far of PCWDEs in the genome, even though cellulose is the least complex of three classes of polysaccharides. These enzymes can be further divided into β-glucosidases, endoglucanases, exoglucanases, lytic polysaccharide monooxygenases (LPMO) and more. β-glucosidases cleave single glucose molecules from the end of oligosaccharides. Endoglucanaese cleave glucans internally while exoglucanases cleave cellobiose residues from the termini of glucose polymers. LPMOs cleave cellulose oxidatively rather than hydrolytically and can function on all parts of the glucan chain. They require copper, molecular oxygen and an electron donor for proper function (18). Below is a table of the gene ID and annotations for each cellulase and LPMO in the *N. crassa* genome. Also included in the right most column is the membership of each enzyme to the glycosyl hydrolase family.

Table 1-1 *N. crassa* cellulases and LPMOs

Gene ID	Annotation	GH Family
NCU00762	β-1,4-endoglucanase EG2 GH 5-1	GH 5
NCU05057	β-1,4-endoglucanase EG-1 GH 7-1	GH 7
NCU04854	β-1,4-endoglucanase GH 7-2	GH 7
NCU04027	β-1,4-endoglucanase GH 7-3	GH 7
NCU05121	β-1,4-endoglucanase GH 45-1	GH 45
NCU05955	β-1,4-endoglucanase GH 74-1	GH 74
NCU00206	cellobiose dehydrogenase	
NCU05923	cellobiose dehydrogenase	
NCU03996	exo-β-1,4-glucanase or cellobiohydrolase Gh6-1	Gh6
NCU09680	exo-β-1,4-glucanase or cellobiohydrolase GH6-2	GH6
NCU07190	exo-β-1,4-glucanase or cellobiohydrolase GH6-3	GH6
NCU07340	cellobiohydrolase-1 CBH-1 GH7-1	GH7
NCU05104	exo-β-1,4-glucanase or cellobiohydrolase GH7-4	GH7
NCU08760	polysaccharide monooxygenase 1 (PMO1)	AA9
NCU03328	polysaccharide monooxygenase 1 (PMO1)	AA9
NCU00836	polysaccharide monooxygenase 1 (PMO1)	AA9
NCU01867	polysaccharide monooxygenase 1 (PMO1)	AA9
NCU02344	polysaccharide monooxygenase 1 (PMO1)	AA9
NCU09764	polysaccharide monooxygenase 1 (PMO1)	AA9
NCU02240	polysaccharide monooxygenase 2 (PMO2)	AA9
NCU02916	polysaccharide monooxygenase 2 (PMO2)	AA9
NCU01050	polysaccharide monooxygenase 2 (PMO2)	AA9

NCU07760	polysaccharide monooxygenase 3 (PMO3)	AA9
NCU03000	polysaccharide monooxygenase 3 (PMO3)	AA9
NCU05969	polysaccharide monooxygenase 3 (PMO3)	AA9
NCU07520	polysaccharide monooxygenase 3 (PMO3)	AA9
NCU07898	polysaccharide monooxygenase 3 (PMO3)	AA9
NCU07974	polysaccharide monooxygenase 3 (PMO3)	AA9

Xylanases

After cellulases, the most abundant group of PCWDEs in *N. crassa* genome are xylanases. These enzymes are responsible for the deconstruction of xylan. Endoxylanases cleave the xylan backbone, while β-xylosidases cleave xylose oligomers in to xylose. *N. crassa* also has a number of acetyl xylan esterases that cleave acetyl moities from the xylan backbone. Below is a table containing the *N. crassa* xylanases, which will be referenced in further chapters.

Table 1-2 Xylanases

Gene ID	Annotation	GH family
NCU05924	β-1,4-endoxylanase GH10-1	GH10
NCU08189	β-1,4-endoxylanase GH10-2	GH10
NCU04997	β-1,4-endoxylanase GH10-3	GH10
NCU07130	β-1,4-endoxylanase GH10-4	GH10
NCU02855	β-1,4-endoxylanase GH11-1	GH11
NCU07225	β-1,4-endoxylanase GH11-2	GH11
NCU09923	extracellular β-xylosidase GH3-7	GH3
NCU00709	extracellular β-xylosidase GH3-8	GH3
NCU04870	acetyl xylan esterase	CE1
NCU04494	acetyl xylan esterase	CE1
NCU00710	acetyl xylan esterase	CE1
NCU05159	acetyl xylan esterase	CE1
NCU09663	acetyl xylan esterase	CE1
NCU09664	acetyl xylan esterase	CE1
NCU03181	acetyl xylan esterase	CE1

Pectinases

N. crassa has a number of pectinases, though significantly fewer than the number of cellulases or xylanases. Included in this list are the enzymes that act on the backbone of pectin, which include galacturonanases and pectate lyases that cleave GalA chains in HG, rhamnogalacturonan lyases that cleave alternating GalA and rhamnose residues in RG1, as well as acetyl esterases that cleave acetyl groups from either type of backbone. Below is a table of the *N. crassa* pectinases.

Table 1-3 Pectinases

Gene ID	Annotation	GH Family
NCU02369	Endogalacturonase	GH28
NCU06961	Exogalacturonase	GH28
NCU06326	pectate lyase	PL1
NCU08176	pectate lyase	PL3
NCU10045	pectin methyl esterase CE8-1	CE8
NCU09976	rhamnogalacturonan acetyl esterase CE12-1	CE12
NCU05598	rhamnogalacturonan lyase ASD-1	PL4

Other hemicellulases/pectinases

N. crassa has a number of other hemicellulases and pectinases predicted in the genome that are not included in previous tables. Many of the sugars present in the plant cell wall are both present in hemicellulose and pectin. Enzymes capable of cleaving these sugars from a polymer can likely act on both hemicellulose and pectin. These include arabinases, galactanases, mannanases, glucuronidases. An additional group of enzymes known as feruloyl esterases removes feruloyl groups abundant in plant polysaccharides. These enzymes are listed in the table below (Table 1-4). Mannanases/mannosidases annotated to cleave linkages other than β-1,4 were omitted due to their likely involvement in protein glycosylation. Ferulates are organic compounds covalently bound to side chains arabinoxylans and other polysaccharides, and also present in lignin.

Table 1-4 Hemicellulases/Pectinases

Gene ID	Annotation	GH Family
NCU09170	α-arabinofuranosidase GH43-4	GH43
NCU05965	α-arabinofuranosidase GH43-7	GH43
NCU02343	α-arabinofuranosidase GH51-1	GH51
NCU09775	α-arabinofuranosidase GH54-1	GH54
NCU00642	extracellular β-galactosidase GH35-1	GH35
NCU04623	extracellular β-galactosidase GH35-2	GH35
NCU06861	glycosyl hydrolase family 43 protein	GH43
NCU01900	intracellular β-xylosidase/arabinosidase GH43-2	GH43
NCU00972	arabinogalactan endo-1,4-β-galactosidase GH53-1	GH53
NCU05882	putative endo-β-1,6-galactanase Gh5-5	Gh5
NCU09702	endo-β-1,6-galactanase GH5-6	GH5
NCU00852	Endoarabinanase GH43-1	GH43
NCU09924	Exoarabinanase	GH93
NCU00937	extracellular β-glucuronidase GH79-1	GH79
NCU09774	feruloyl esterase	AA12
NCU09491	feruloyl esterase FEA-1	
NCU07351	α-glucuronidase	GH67
NCU06143	α-glucuronidase	GH67

NCU04885	α-xylosidase	GH31
NCU08412	β-1,4-endomannanase GH5-7	GH5
NCU00985	extracellular β-mannosidase GH2-4	GH2
NCU00890	intracellular β-mannosidase Gh2-1	Gh2

Starch degrading enzymes

N. crassa has a small suite of starch degrading enzymes. The majority of these proteins are α-glucosidases, which cleave α linkages between glucose monomers, and include the α-amylases listed below(Table 1-5). *N. crassa* also has a single trehalase, and one starch active LPMO (19). However, many of these enzymes may not be active on external starch sources, as the protein sequences lack any signal peptide as determined by SignalP (20). Included in Table 1-5 is an additional column indicating whether the protein is likely intracellular or extracellular, depending on presence or absence of a predicted signal peptide.

Table 1-5 Starch active enzymes

Gene ID	Annotation	GH Family	cell loc.
NCU08131	α-amylase	GH13	extracellular
NCU09805	α-amylase	GH13	extracellular
NCU05873	α-amylase	GH13	intracellular
NCU05429	α-amylase	GH13	intracellular
NCU06523	intracellular α-glucosidase GH13-4	GH13	intracellular
NCU07860	intracellular α-glucosidase GH13-5	GH13	intracellular
NCU03098	intracellular α-glucosidase GH15-1	GH15	intracellular
NCU00743	amylo-α-1,6-glucosidase GH13-7	GH13	intracellular
NCU01517	extracellular α-glucosidase gla-1		intracellular
NCU02583	extracellular α-glucosidase gla-2		extracellular
NCU09281	extracellular α-glucosidase GH31-1	GH31	extracellular
NCU04203	extracellular α-glucosidase GH31-2	GH31	extracellular
NCU04674	extracellular α-glucosidase GH31-3	GH31	extracellular
NCU04221	neutral trehalase	GH37	intracellular
NCU08746	Starch active LPMO	AA13	extracellular
NCU00943	trehalase	GH37	extracellular

1.5. Post translational modifications to PCWDEs

Post-translational modifications (PTMs) are modifications to a peptide chain or protein that occur after the mRNA transcript has been translated into protein. These modifications affect protein properties and functions. Some common examples of PTMs include the attachment of functional groups to amino acid side chains, disulfide bond formation, and phosphorylation. PCWDEs are generally secreted proteins that pass through the secretory pathway, the process during which many post-translational modifications occur. PCWDEs have several PTMs that are important for their function.

Glycosylation is the most studied PTM in PCWDEs due to its potential to affect enzyme activity and binding. Glycosylation is the covalent attachment of carbohydrates to amino acid side chains of a protein. Research has been conducted to better understand the role of glycosylation in cellulases and cellulase activity. So far, data suggests that glycosylation can affect substrate binding, protein stability, resistance to proteases and modification of protein hydrophobicity (21). Understanding glycosylation is especially important for heterologous expression of PCWDEs, as host organisms may have unique glycosylation machinery, which might result in unexpected properties of the PCWDE of interest.

The PTM discussed in this dissertation is pyroglutamate modification. Like glycosylation, it has been shown previously to have a substantial affect on protein stability and activity (22). Chapter 2 explores the formation, function, and prevalence of pyroglutamate in fungal proteins.

1.6. **Fungal Metabolic Transcription factors**

DNA binding is an important function for many proteins with a wide range of essential functions. Gene regulation is one such function, and is governed by transcription factors (TFs) that activate and repress expression of a given gene set at an appropriate time. Fungi possess TFs from number of different families. These families are organized by very specific DNA-binding motifs. Different families include the basic helix-loop-helix, helix-turn-helix, MADS box, basic region leucine zipper (bZIP), and zinc(II) coordinating TFs, GATA, C2H2, and Zn2Cys6 (23, 24). Fungi have a large set of Zn2Cys6 TFs, the function of which is unknown. The Zn2Cys6 family also contains all of the most critical TFs involved in carbon metabolism and PCWDE production that have been described to date. This section is meant to familiarize the reader with important Zn2Cys6 TFs identified with roles in carbon metabolism, as well as crucial TFs involved with nitrogen, sulfur, and phosphate metabolism, all of which are further discussed in upcoming chapters.

***xlr-1/xyr-1/xlnR* (Zn2Cys6)**

XLR-1 is the major TF responsible for regulating cellulose and hemicellulose degradation in both *Trichoderma reesei* and *Aspergillus niger* (25, 26). However, in other Ascomycota like *N. crassa, Aspergillus nidulans,* and *Penicillium oxalicum* XLR-1 (XlnR) function is partially or fully decoupled from cellulose degradation (27–30). In *N. crassa,* XLR-1 has been shown to regulate expression of xylanases, as well as being critical for expression of two xylose metabolic genes, xylose reductase (NCU00891) and xylitol dehydrogenase (NCU08384) (30, 31).

***clr-1/clr-2* (Zn2Cys6)**

CLR-1 and CLR-2 are the major cellulase regulators in *N. crassa* (27, 31). CLR-1 is activated in the presence of cellulose, and upregulates expression of *clr-2*, the major transcriptional activator of cellulases. Interestingly, CLR-2 seems to be the only described carbon metabolism regulator in Ascomycota that is active in its native state. Mis-expression or overexpression of CLR-2 induces the expression of

cellulases even under carbon catabolite repression conditions (CCR). This seems to only be true for CLR-2 in *N. crassa*, as this phenotype was not observed in the homologs of *clr-2* in *P. oxalicum* or *A. nidulans* (28, 32). Over-expression of just *clr-1* had no affect of cellulase expression under CCR conditions.

***manR* (Zn2Cys6)**

ManR a homolog of *clr-2*. Re-discovered in *Aspergillus oryzae* in 2012, MANR has been shown to regulate expression of a β-mannanase and other enzymes involved in depolymerization of galactomannans and glucomannans (33). It was also shown to regulate cellulose utilization in *A. niger* (34) This was also was shown in *N. crassa.*

***pdr-1/RhaR* (Zn2Cys6)**

PDR-1 was simultaneously discovered in *A. niger* and *N. crassa,* and is the major rhamnose response regulator in both fungi. It is critical for full pectin utilization, rhamnose utilization, and has a minor role in GalA utilization. In *A. niger,* RhaR has a lesser role in GalA utilization (35, 36).

***GaaR/pdr-2* (Zn2Cys6)**

GaaR was originally discovered in *Botrytis cinerea* in 2015 when transcriptome data revealed it as a GalA inducible gene. Deletion of GaaR results in the inability of *B. cinerea* to utilize GalA (37). GaaR was further investigated in *A. niger* and was shown to regulate expression of a number of pectinases in addition to regulating GalA metabolism (38). GaaR was further shown to act in concert with a repressor protein GAAX that is inactivated in the presence of GalA (39). We have named the homolog of GaaR in *N. crassa* as *pdr-2*. The discovery and description of *pdr-2* is discussed in chapter 3.

***araR* (Zn2Cys6)**

AraR, discovered in *A. niger* in 2011, controls extracellular L-arabinose releasing enzymes, as well as arabinose catabolic pathway but only within the Eurotiales (40). It appears to originate from gene duplication of *xlnR*.

***ara-1* (Zn2Cys6)**

Ara-1 was discovered in 2016 in *Pyricularia oryzae* (*Magnaporthe oryzae).* It was found through characterization of candidate TFs derived from RNAseq data comparing response between fructose and arabinose (41). It was shown to regulate key arabinose catabolism genes and an extracellular arabinofuranosidase Discovery and characterization of this TF in *N. crassa* is further discussed in Chapter 3.

***cre-1/cre1/creA* (Zn2Cys6)**

This TF is the most well known negative regulator of PCWDEs, and is integral to fungal CCR. Originally identified *A. nidulans* four decades earlier (42), it has since been characterized in *N. crassa, T. reesei, A. niger* and many other members of Ascomycota. CRE-1 functions by repressing gene expression PCWDEs when more favorable sugars like glucose are in abundance. In this way, the fungus does not

expend unnecessary energy to make PCWDEs. In *N. crassa* and *P. oxalicum,* it is thought that cellulolytic genes as well as *clr-1/clr-2* are regulated by CRE-1 (32, 43). CRE-1 mediated CCR is explored further in Chapter 5 and 6.

vib-1

VIB-1 plays a role in heterokaryon incompatibility, phosphatase expression, protease expression and more (44–47). More recently, *vib-1* deletion strains were shown to be unable to grow on Avicel (crystalline cellulose) and to produce cellulases (48). Over-expression of *vib-1* in *T. reesei* confirms role in cellulase production (49). We will further discuss this TF in this dissertation.

col-26

Deletion of this *col-26* results in colonial growth, a phenotype associated with lack of colony expansion. It has been shown to be involved in glucose sensing, carbon catabolite repression, regulation of starch metabolism, and integration of carbon and nitrogen metabolism (48, 50).

Nitrogen metabolism: *nit-2* (GATA), *nit-4* (Zn2Cys6) and *NMR*

NIT-2 and NIT-4 are the two major transcription factors responsible for regulating nitrogen metabolism *N. crassa*. Inorganic nitrate is a nitrogen source for fungi like *N. crassa* and *Aspergillus*, but is not utilized in the presence of a preferred carbon source like ammonia, glutamine and glutamate. These two TFs, along with the repressor NMR, work together to regulate genes involved in nitrate utilization (51). Their regulation of *nit-3* (nitrate reductase) is the canonical model for how the two TFs interact. In the presence of nitrate, NIT-4 is activated and can cooperatively bind the promoter of *nit*-3 with NIT-2 to promote transcription. When nitrate is absent, NIT-4 is inactive and the pair are unable to induce expression of *nit-3*. If nitrate is present along with ammonia/glutamine/glutamate, NMR is active and negatively regulates NIT-2 so that the pair cannot induce expression of *nit-3* (51). Modern studies have yet confirm the full regulon of these transcription factors.

Phosphate metabolism and *nuc-1* (Basic Helix Loop Helix)

NUC-1 is the major phosphate metabolism regulator. It is active in absence of phosphate. In its active state, NUC-1 induces expression of a number of phosphate acquisition genes, notably several alkaline phosphatases and inorganic phosphate transporters (52). Modern studies have yet explore the full regulon of NUC-1 and phosphate regulation.

Sulfur metabolism *cys-3* (bZip), *scon-1* and *scon-2*

CYS-3 is the major positive regulator of sulfur acquisition genes in *N. crassa*. Under sulfur repression, SCON1 and 2 are both involved in preventing *cys-3* expression. Research has shown that the presence of CYS-3 itself is enough to promote expression of its targets, indicating that CYS-3 need not be activated by any process, nor directly antagonized by binding of another molecule (53). Interestingly,

the *scon-2* promoter has several CYS-3 binding sites, indicating that a negative feedback loop regulates *cys-3* expression (54).

Chapter 2. **Pyroglutamate modification and its role in CBH-1 and other fungal secreted proteins**

2.1. **Abstract**

A post-translational N-terminal pyroglutamate modification observed in certain PCWDEs occurs when N-terminal glutamine or glutamate is cyclized to form a five-membered ring. This modification has been shown to confer resistance to thermal denaturation for CBH-1 and EG-1 cellulases. In mammalian cells, the formation of pyroglutamate is catalyzed by glutaminyl cyclases. Using the model filamentous fungus *Neurospora crassa,* we identified two genes (*qc-1* and *qc*-2) that encode proteins homologous to mammalian glutaminyl cyclases. We show that *qc-1* and *qc-2* are essential for catalyzing the formation of an N-terminal pyroglutamate on CBH-1 and GH5-1. CBH-1 and GH5-1 produced in a *Δqc-1Δqc-2* mutant, and thus lacking the N-terminal pyroglutamate modification, showed greater sensitivity to thermal denaturation, and for GH5-1, susceptibility to proteolytic cleavage. QC-1 and QC-2 are ER-localized proteins. The pyroglutamate modification is predicted to occur in a number of additional fungal proteins that have diverse functions. The identification of glutaminyl cyclases in fungi may have implications for production of lignocellulolytic enzymes, heterologous expression and biotechnological applications revolving around protein stability.

2.2. **Introduction**

Research on the characterization of lignocellulolytic enzymes in fungi is extensive and crystallography studies have revealed important post-translational modifications. For example, the crystal structure of Cbh1, a cellobiohydrolase belonging to the glycosyl hydrolase (GH) 7 family revealed an N-terminal five-membered cyclic ring structure known as pyroglutamate (pGlu) (Figure 2-1A). N-terminal pGlu modification has been reported in a wide variety of proteins found in plants, animals and bacteria and is catalyzed by glutaminyl cyclases (QC)(55–57). In mammalian systems, a secreted QC catalyzes pGlu-modified amyloid peptides, which are associated with increasing the risk of Alzheimer's disease (58). The importance of pGlu modification was first demonstrated for its role in function and stability of peptide hormones and signaling molecules (59–61), snake venom peptides and certain ribonucleases (62, 63). Less is known about the role of functionally convergent QCs in plants but pGlu modification has been detected in a small number of proteins involved in wounding and pathogen response (56).

Although pGlu modification is evident in crystal structures of a number of lignocellulolytic enzymes in filamentous fungi, the biological role of this modification is unclear. However, two recent studies showed that heterologous

production of a cellobiohydrolase-1 (CBH1) and an endoglucanase-1 (EG1) from a thermotolerant filamentous fungus *Talaromyces emersonii* in *Saccharomyces cerevisiae* showed decreased thermal stability (22, 64)(Figure 2-1B). The thermal stability of *T. emersonii* CBH1 (TeCBH1) and EG1 purified from *S. cerevisiae* was rescued following treatment with a human QC enzyme.

Here, we explore pGlu formation and function using the lignocellulolytic fungus *Neurospora crassa* as a model system and identify two functionally redundant ER-localized enzymes that are essential for pGlu modification of a GH7 enzyme (TeCBH1 and native CBH-1) and a GH5 enzyme (endoglucanase GH5-1). We show that the pGlu modification of CBH-1 and GH5-1 enhances thermostability and for GH5-1, resistance to N-terminal proteolytic cleavage. The development of an *N. crassa* system devoid of pGlu modification will allow further investigations into the diverse role of pGlu modifications on a variety of proteins that traffic through the secretory pathway in filamentous fungi and may enable the development of tools for biotechnological applications.

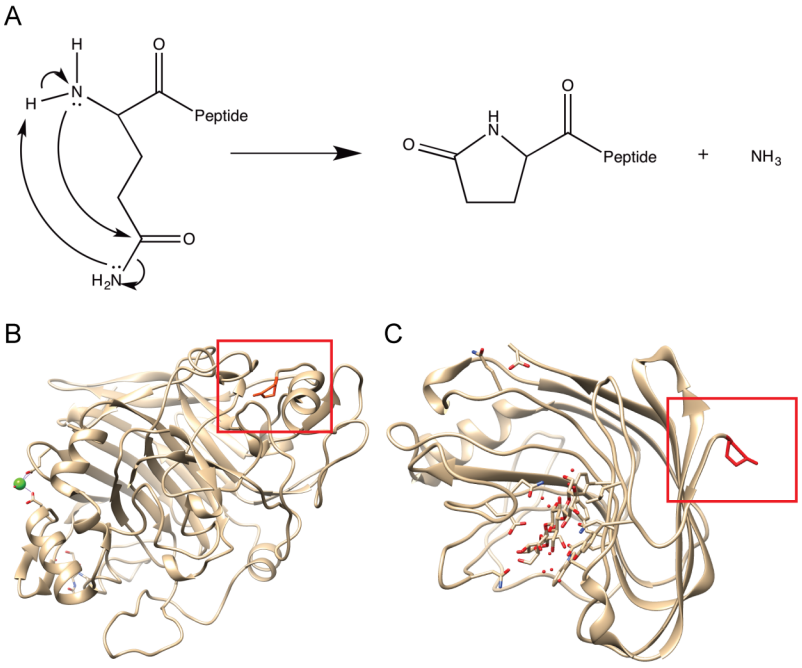


Figure 2-1 Pyroglutamate formation and presence and function in fungal hydrolytic enzymes. (A) Mechanism of pyroglutamate formation from N-terminal

glutamine. Conversion. of glutamine to pGlu releases one ammonia molecule. (B) Crystal structure of CBH-1 (GH7 family) from *T. reesei* shows pGlu-modified N-terminus with pGlu distinctly tucked into pocket within the protein (65). (C) Crystal structure of Cel12A (GH12 family) from *H. grisea* provides an example of pGlu-modified N-terminus typical of non-GH7 family proteins (66). Crystal structure images generated with Chimera 1.10.2 (67) from structures fetched from RCSB Protein Data Bank (PDB ID 1CEL, 1UU4).

2.3. **Materials and methods**

Phylogeny. Protein sequences were obtained by BLAST using NCU09018 (QC-1) and NCU11249 (QC-2) as protein search queries. Sequences were aligned using MAFFT version 7 (68). The alignment was used to construct a maximum likelihood phylogeny using RAxML (69). FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used for visualization.

Strain Construction. Cassettes consisting of a hyrogomycin B resistance marker flanked by 3' and 5' regions of NCU11249 (*qc-2)* or NCU09018 (*qc-1)* were obtained from the Dunlap Lab (Geisel School of Medicine, Dartmouth). Cassettes were transformed into a Δ*mus-52* strain as described previously (70). Hygromycin B-resistant strains were verified to the QC deletion through PCR analysis (Fig S2). Strains were crossed back to *N. crassa* FGSC 2489 and *qc-1* or *qc-2* deletion strains lacking Δ*mus-52* were identified; strains are available at the Fungal Genetics Stock Center (FGSC). The Δ*qc-1*Δ*qc-2* strain was generated by crossing Δ*qc-1* and Δ*qc-2* strains and screening ascospore progeny for the *qc-1* and *qc-2* deletions by PCR (Fig S2C). For complementation, *qc-1* and *qc-2* were amplified from genomic DNA with additional restriction sites and fused to super folder GFP (71). Forward primer GAAAGGATCCATGACACGACGCAACCGCTT and reverse primer GATTAATTAAA-GGCCCTTTCCCAGCA for *qc-1* and forward primer GAAAGGATCCATGGTGACGATG-GCACG and reverse primer GATTAATTAAGAGCTCGTCTTTCTCCATGCTC for *qc-2* included restriction sites for BamH1 and PacI for cloning. Each insert was cloned into the pCSR1::GPD vector (22, 72), transformed into a Δ*qc-1* Δ*qc-2* strain and selected for cyclosporin resistance as described.

The GH5-1 ORF (NCU00762) was PCR amplified from *N. crassa* genomic DNA with an additional 6XHis sequence and restriction sites and inserted into the pCSR1::GPD vector (22). Forward primer AATCAAAGCGGCCGCATGAAGGCTACGAT-TCTTGCCA and reverse primer AAATTAATTAATTAATGATGATGATGATGATGAGG-GGTATAGGTCTTGAGAAGG included restriction sites for NotI and PacI restriction enzymes. The TeCBH-1 in pCSR1::GPD vector (22) and GH5-1 in the pCSR1::GPD were transformed into *N. crassa* as described previously (72). Homokaryotic transformants were obtained as described above. The Sec61::mCherry strain was described previously (73).

Fluorescence Microscopy. Confocal microscopy was performed as described (73) using a Leica SD6000 microscope with a 100x1.4 NA oil-immersion objective

equipped with a Yokogawa CSU-X1 spinning disc head and a 488-nm or 561-nm laser controlled by Metamorph software. ImageJ software was used for false color and overlaying images.

CBH-1 and GH5-1 Purification. For TeCel7A and GH5-1 purification, conidia from a homokaryotic *csr-1::TeCel7A* strain or a *csr-1::GH5-1-6xHis* strain were harvested and used to inoculate 1L of Vogel's minimal medium (VMM) in a 2L flasks at a concentration of 10^6 conidia/ml. Cultures were grown at 25˚C 200 rpm shaking for 3 days. The cultures were filtered through a glass fiber filter and Corning 22 μm PES bottle top filter. The supernatant was concentrated 10 fold using a Pellicon 5KD filter cassette and proteins were precipitated using 45 g ammonium sulfate. Proteins were pelleted and re-suspended in 20mM Tris-HCL and desalted using HiPrep 26/10 desalting column (GE Healthcare Life Sciences). TeCBH-1 was separated from other extracellular proteins using the anion exchange column MonoQ 10/100 (GE Healthcare Life Sciences), while a GE Healthcare Life Sciences 1ml His-Trap HP column (GE Healthcare Life Sciences) was used was used in tandem with MonoQ 10/100 column to separate GH5-1 from other extracellular proteins.

MuLac Assay for CBH-1. Purified CBH-1, MuLac (4-methylumbelliferyl β-D-lactopyranoside; Sigma) at 1mM final concentration was incubated with CBH-1 enzyme at 15ug/ml in 0.05M sodium acetate buffer pH 5 in 50ul volumes in VWR 12 well 0.2ml PCR strip tubes. Strip tubes were incubated for 15 min across temperatures using Applied Biosystems Veriti 96-well thermocycler in veriflex mode. After 15 min, temperature was raised to 95˚C for 5 min to deactivate the protein. Reactants were transferred to corning half well 96 clear bottom plates, and fluorescence was measured at 445nm after excitation with 365nm.
For TeCBH-1 expressed in *N. crassa* supernatant assay, MuLac (1mM final concentration) was incubated with 25ul of filtered culture supernatant for a final 50ul volume in 0.05M sodium acetate buffer pH 5. Strip tubes were incubated using Applied Biosystems Veriti 96-well thermocycler in verflex mode for 1 hr. The 445-nm emission/365-nm excitation was measured post incubation.

Filter Paper Assay. Purified GH5-1 was assayed using a modified filter paper assay (74). Folded 1-cm squares of Whatman 1 filter paper were incubated with purified GH5-1 in 0.05M citrate buffer pH 4.8 at a concentration of 7.5ug/ml. The strips were placed in tubes and incubated in an Applied Biosystems Veriti 96-well thermocycler in veriflex mode holding every two wells at varying temperatures, 6 temperatures per 12 well strip tube. Wells were held between 20-45˚C or 50-75˚C at 5˚C increments for 24h. After incubation, 100ul of 3,5-dinitrosalicylic acid reagent was added to assay for oligosaccharide reducing ends. Samples were diluted and measured for absorbance at 540nm.

Differential Scanning Calorimetry. Purified CBH-1 at ~500ng/ul in a 20mM Tris-HCL pH 8.5 buffer was loaded into the sample chamber Nano DSC (TA Instruments) and assessed for stability by differential scanning calorimetry according to (22).

Liquid Chromatography-Mass Spectrometry. Trypsin-digested protein samples were analyzed using a Thermo-Dionex UltiMate3000 RSLCnano liquid chromatograph that was connected in-line with an LTQ-Orbitrap-XL mass spectrometer equipped with a nanoelectrospray ionization (nanoESI) source (Thermo Fisher Scientific, Waltham, MA). The LC was equipped with a C18 analytical column (Acclaim® PepMap 100, 150 mm length × 0.075 mm inner diameter, 3 μm particles, 100 Å pores, Thermo) and a 1 μL sample loop. Acetonitrile (Fisher Optima grade, 99.9%), formic acid (1 mL ampules, 99+%, Thermo Pierce), and water purified to a resistivity of 18.2 MΩ·cm (at 25 °C) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile phase solvents. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). The elution program consisted of isocratic flow at 2% B for 4 min, a linear gradient to 30% B over 38 min, isocratic flow at 95% B for 6 min, and isocratic flow at 2% B for 12 min, at a flow rate of 300 nL/min.

Full-scan mass spectra were acquired in the positive ion mode over the range m/z = 350 to 1800 using the Orbitrap mass analyzer, in profile format, with a mass resolution setting of 30,000 (at m/z = 400, measured at full width at half-maximum peak height, FWHM). In the data-dependent mode, the eight most intense ions exceeding an intensity threshold of 50,000 counts were selected from each full-scan mass spectrum for tandem mass spectrometry (MS/MS) analysis using collision-induced dissociation (CID). MS/MS spectra were acquired using the linear ion trap, in centroid format, with the following parameters: isolation width 3 m/z units, normalized collision energy 30%, default charge state 3+, activation Q 0.25, and activation time 30 ms. Real-time charge state screening was enabled to exclude unassigned and 1+ charge states from MS/MS analysis. Real-time dynamic exclusion was enabled to preclude re-selection of previously analyzed precursor ions, with the following parameters: repeat count 3, repeat duration 10 s, exclusion list size 500, exclusion duration 90 s, and exclusion mass width 20 ppm. Data acquisition was controlled using Xcalibur software (version 2.0.7, Thermo). Raw data were searched against the amino acid sequence of *N. crassa* CBH-1 (NCU07340) using Proteome Discoverer software (version 1.3, SEQUEST algorithm, Thermo) for tryptic peptides with up to three missed cleavages, and carbamidomethylcysteine, methionine sulfoxide, and N-terminal pyroglutamate as variable post-translational modifications. Peptide identifications were validated by manual inspection of the MS/MS spectra, i.e., to check for the presence of y-type and b-type fragment ions[S1] that identify the peptide sequences. Data acquisition and integration of extracted ion chromatograms of peptide ions were performed using Xcalibur software (version 2.0.7, Thermo). MS/MS spectra are annotated using the nomenclature of Roepstorff and Fohlman (75).

GH5-1 Immunoprecipitation and Western blot analyses. Polyclonal rabbit antibodies were obtained using the synthetic peptide DPENKIVYEMHQYLDSD from GH5-1 (Pierce Biotechnology). GH5-1-6xHis was immunopreciptated (IP) from cell lysate using Thermo Fisher Dynabeads Protein G bound to anti-GH5-1 antibody.

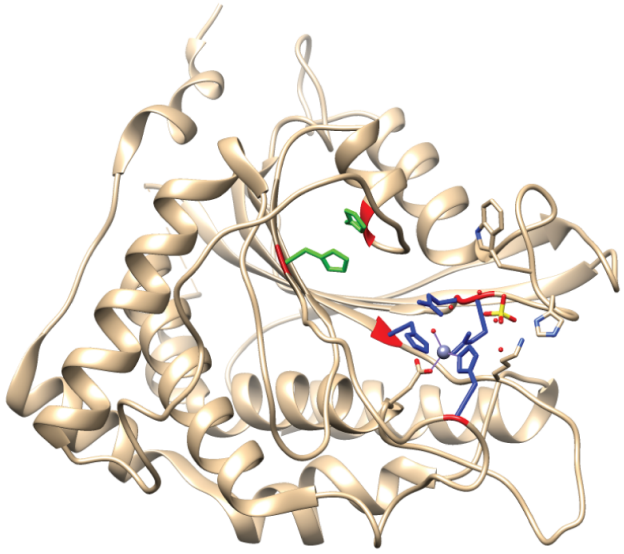
Beads were washed and incubated for 1 hour at 4˚C with cell lysate, after which beads were washed with citrate-phosphate buffer (pH 5.7). Proteins were eluted with 100mM citrate buffer (pH 2.4). Immuno-preciptated proteins were subjected to SDS PAGE and transferred to nitrocellulose membrane for blocking and blotting. Membrane was developed using Thermo Fisher Super Signal West Pico Chemiluminescent substrate.

2.4. **Results**

2.4.1. ***N. crassa has two QCs homologous and conserved with Human QC***

We hypothesized that fungi have uncharacterized QC enzymes that catalyze pGlu formation in lignocellulolytic enzymes. We searched the *N. crassa* genome for homologs to the human pituitary QC (*QPCT* Gene ID 25797) biochemically shown to catalyze pGlu formation (76). We identified two genes, NCU09018 and NCU11249, that encode predicted proteins with significant similarity to human QC (~25%-30% amino acid identity). The NCU09018 and NCU11249 proteins show conservation of four catalytic residues (H140, E201, E202, H330) and two substrate-binding residues (H309, H317) that are important for function of human QC enzymes (Figure 2-1) (76). We named NCU09018 and NCU11249 *qc-1* and *qc-2*, respectively.

A



B

```
Homo_sapiens_QC  ------------------MAGGRHRR-----VVGTLHLLLLVAALPWAS  26
   NCU11249_QC2  MVTMARHIGLDPNNPQ--WHRSSTSLS-------RHLLSALLL-TTTIT-  39
   NCU09018_QC1  ---MTRRNRFATRSPQPTITAGTKSLRPTTMSALPTLLSLLFAAVVLFV-  46

Homo_sapiens_QC  RGVSPSASAWPEEKNYHQPAILNSSALRQIAEGTS----ISE--MWQNDL  70
   NCU11249_QC2  ---TGVH-AY---------TPLSTCTLQHLTFGPSSDFDIHNGFGSSSLL  76
   NCU09018_QC1  ---APSL-AY---------QPLSDTQLKAIPSPLNSDFDIKT----GALL  79

Homo_sapiens_QC  QPLLIERYPGSPGSYAARQHIMQRIQRLQADWVLEIDTFLSQTPYGY---  117
   NCU11249_QC2  APILIPRVPGTEGSRLVQQHFVDFFSSQLPDWTLEWQNSTSTTPATGSQL  126
   NCU09018_QC1  APILIPRVPGTPGQAKVQKHFVDFFSRELPEWDISWQNSTATTPLSGKKQ  129

Homo_sapiens_QC  RSFSNIISTLN--------PTAKRHIVLACHYDSKYFSHWNNRVFVGATD  159
   NCU11249_QC2  IPFANLILRRDPPWAKAG---NVKRHTLAAHYDSLF----RPEGFIGAVD  169
   NCU09018_QC1  IPFQNLIFRREPPWTKERGPGRAALLTLVAHYDSKI----SPEGFIGATD  175

Homo_sapiens_QC  SAVPCAMMLELARALDKKLSL-------------------------KTV  184
   NCU11249_QC2  SAAPCAILMAVARAVDGALGRRWEGVMAAKEKREGGGERDAGDGLEDEEG  219
   NCU09018_QC1  SAAPCAVLMHVARTVEGYLKKVYEEGVS--------------GGLGKEGR  211

Homo_sapiens_QC  SDSKPDLSLQLIFFDGEEAFLHWSPQDSLYGSRHLAAKMASTFHPPGARG  234
   NCU11249_QC2  GEEEEEKGVQIVLFDGEEAWERWTNTDSTYGSRALAEAWQSSFYEASSTH  269
   NCU09018_QC1  EDPKREVGVQILLLDGEEAFKEWTDTDSLYGARSLSEEWENTFYPALSRF  261

Homo_sapiens_QC  TSQLHGMDLLVLLDLIGAPNPTFPNFFPNSARWFERLQAIEHELHELGLL  284
   NCU11249_QC2  SNRLESISLLVLLDLLGAGNPRIPSYFWDTHGAYKDLAKIETRLRKLGVL  319
   NCU09018_QC1  ANPIRQIDLFVLLDLLGSADPGVPSYFQTTHWAYKNMATVESRMRALGLL  311

Homo_sapiens_QC  KDHSLEG-----RYFQNYSYGGVIQDDHIPFLRRGVPVLHLIFSPFPEVW  329
   NCU11249_QC2  ETAPASPFLPDSEKPYNRFTRGYIQDDHVPFMERGVKVLHIIETPFPPVW  369
   NCU09018_QC1  ESKPKDPFLPEAGKLKEHFGRAYVGDDHQPFMAKGAPVLHMIETPFPHVW  361

Homo_sapiens_QC  HTMDDNEPNLDESTIDNLNKILQVFVLEYLHL------------------  361
   NCU11249_QC2  HTMDDDGEHLDLPTVRDWAKIMTVFVAEWMDLDGVLSQESCAAQKEKGSG  419
   NCU09018_QC1  HKIEDDGEHLDLPTVRDWARIVTAFTIEYLEATTAEAGAVGGGKEGKGED  411

Homo_sapiens_QC  ------------------------                           361
   NCU11249_QC2  SME---------KDEL---------                           426
   NCU09018_QC1  AAAAAKQGTSDGEEKAGEGAGKGP                           435
```

Figure 2-2 Conservation between *N. crassa* QC-1/QC-2 and Human QC. *(A) Crystal structure of human glutaminyl cyclase* (77) *showing functionally critical conserved residues with predicted glutaminyl cyclases from N. crassa (QC-1 and QC-2) (PDB ID 2AFM). Highlighted in blue are catalytic residues H140, E201, E202 and H330. Highlighted in green are substrate binding residues H309 and H317. (B) Multiple sequence alignment of amino acid sequences from human QC and N. crassa QC-1 and QC-2. Identical amino acids are highlighted in red and similar amino acids are highlighted in pink. Amino acids boxed in blue and green are the same catalytic and substrate-binding residues as described above.*

Using *qc-1* and *qc-2* amino acid sequences as queries, we built a maximum likelihood phylogeny of fungal QC proteins (Figure 2-3). The genomes of the majority of ascomycete fungi have one predicted gene with high similarity to *N. crassa qc-1/qc-2.* Homologs to *qc-1/qc-2* were identified in other major lineages in fungi, including the Chytridiomycota, the Zygomycota, the Glomeromycota and the Basidiomycota. The phylogeny additionally showed that species with two QC genes is the derived state of a small subclade of species within the Sordariomycetes (Figure 2-3).

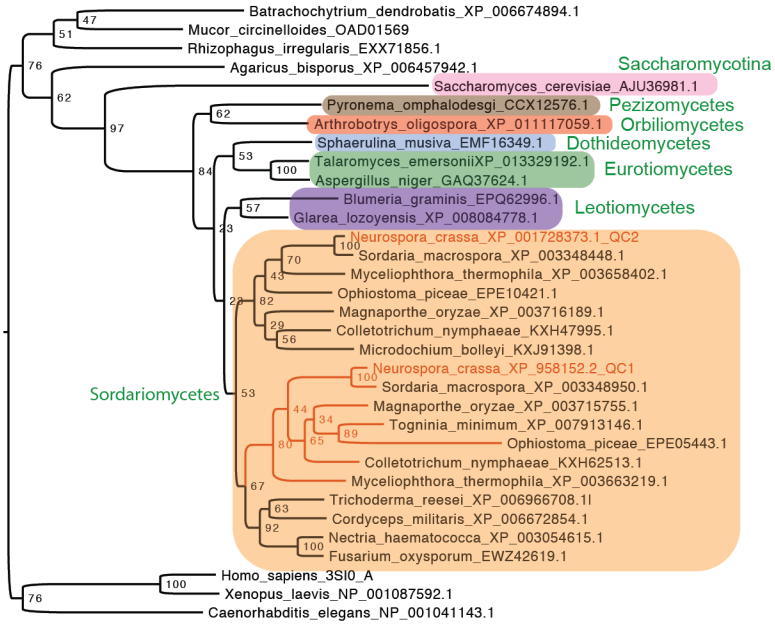


Figure 2-3 QC phylogeny. *Maximum likelihood tree built from homologs of qc-1 identified in fungal genomes using animals as the outgroup. Bootstrap values generated from 100 replicates are labeled at each node. Tree branches in orange highlight species with secondary copy of QC and represent a subclade within Sordariomycetes that contains N. crassa and closely related species. Representatives from various classes within Ascomycota are highlighted in colored boxes and labeled in green.*

2.4.2. ***Deletion of the two predicted glutaminyl cyclase genes (qc-1 and qc-2) in N. crassa results in thermal instability of CBH-1***

To determine whether *qc-1* and/or *qc-2* are important for pGlu formation in CBH-1, we constructed strains carrying deletions of *qc-1 (Δqc-1)*, *qc-2 (Δqc-2)* and both QCs *(Δqc-1Δqc-2).* The *T. emersonii cbh-1 (TeCBH1)* regulated by the glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter (22) was introduced into the wild type (WT), *Δqc-1, Δqc2* and *Δqc-1Δqc-2* strains. Activity of TeCBH1 from culture supernatants was assayed from 35˚C to 60˚C using a MuLac assay for enzyme activity (Fig. 3A) (S1 Material and Methods). TeCBH1 from WT cells and

single *Δqc-1* or *Δqc-2* mutants displayed high activity (Figure 2-4A). In contrast, TeCBH1 from the Δ*qc-1*Δ*qc-2* mutant showed reduced activity at high temperatures, an effect especially noticeable between 60˚C and 65˚C (Figure 2-4A). These results recapitulated previous data demonstrating a difference in activity in pGlu-modified versus unmodified TeCBH1 at different temperatures (22). Furthermore, these data suggested that the *N. crassa* QC enzymes are functionally redundant.

To determine if the role of *qc-1* and *qc-2* in the thermotolerance of heterologously expressed TeCBH1 was unique, we investigated the thermotolerance of native CBH-1 (NCU07340) in WT cells versus the *Δqc-1Δqc-2* mutant. Similar to the results with TeCBH1, native CBH-1 purified from the Δ*qc-1*Δ*qc-2* mutant showed low enzymatic activity between 25˚C and 40˚C (Figure 2-4B), while CBH-1 purified from WT cells showed an increase in enzyme activity between 25˚C and 40˚C. To confirm that the effect on thermostability of CBH-1 was dependent upon functional *qc-1* and *qc-2,* the *Δqc-1Δqc-2* strain was transformed with either *qc-1* or *qc-2* constructs epitope-tagged with in-frame superfolder GFP (*qc-1-sfGFP* or *qc-2-sfGFP*). CBH-1 purified from the *Δqc-1Δqc-2* mutant bearing either *qc-1-sfGFP* or *qc-2-sfGFP* had a similar activity curve to CBH-1 isolated from WT cells, indicating that QC function in *N. crassa* is redundant under these experimental conditions. Importantly, both QC-1 and QC-2 were essential for thermostability of TeCBH1 and CBH-1 in *N. crassa*.

To further define the difference in thermal stability of native CBH-1, we performed differential scanning calorimetry to determine the melting temperatures (T_m) of purified CBH-1 from WT cells versus CBH-1 purified from *Δqc-1Δqc-2* mutant cells. CBH-1 purified from WT cells showed a melt peak at ~48˚C, while CBH-1 purified from the *Δqc-1Δqc-2* cells showed a melt peak at ~39˚C (Figure 2-4C). Consistent with previous results, the melting curves of CBH-1 purified from the *Δqc-1Δqc-2* mutant bearing either *qc-1-sfGFP* or *qc-2-sfGFP* were very similar to that of CBH-1 purified from WT cells (Figure 2-4D).

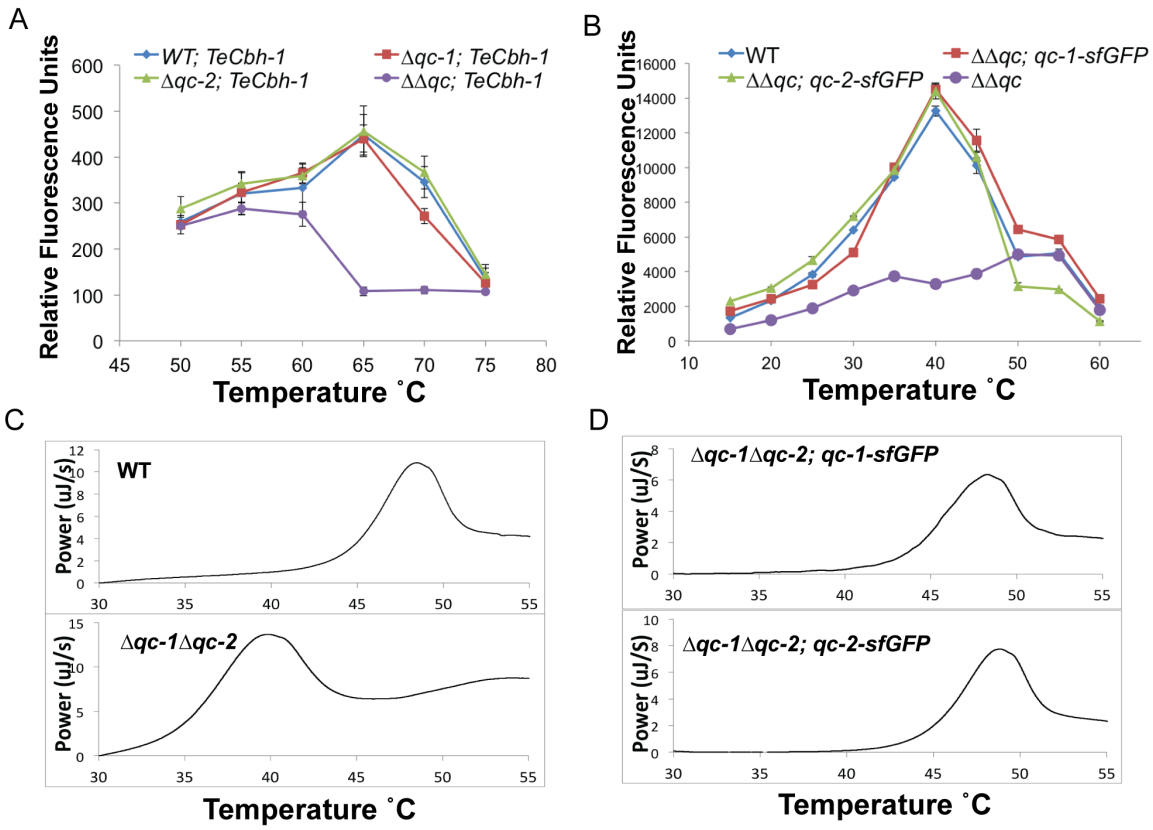


Figure 2-4 TeCBH1/NcCBH1 activity and melting temperatures in the *Δqc* strains. (A) MuLac activity between 50˚C and 75˚C of supernatants from WT, *Δqc-1, Δqc-2* and *Δqc-1Δqc-2* strains expressing TeCBH1. (B) MuLac activity between 15˚C and 60˚C of purified CBH-1 from WT or *Δqc-1Δqc-2* cells or from the *Δqc-1Δqc-2* mutant containing *qc-1-sfGFP* or *qc-2-sfGFP*. Error bars in A and B represent standard deviation from quadruplicates (C, D) Melting temperature of CBH-1 purified from WT, *Δqc-1Δqc-2* or *Δqc-1Δqc-2* rescued cells as measured by differential scanning calorimetry.

2.4.3. *Mass spectrometry of CBH-1 purified from WT cells shows the presence of an N-terminal pyroglutamate modification*

Previous studies describing pGlu function in CBH-1 did not verify the presence or absence of the N-terminal pGlu modification (22, 64). To address this question directly, we conducted liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of trypsin-digested CBH-1 purified from WT cells versus that from the Δ*qc-1*Δ*qc-2* mutant; the N-terminal CBH-1 peptide resulting from trypsin digestion is QAVCSLTAETHPSLNWSK. Formation of pyroglutamate from an N-terminal glutamine residue results in a decrease in exact (monoisotopic) molecular mass of 17.0265 Da, due to elimination of a molecule of ammonia. The difference in molecular mass between N-terminal peptides with and without pyroglutamate was discerned in precursor ions measured in full-scan mass spectra (Figure 2-5A).

MS/MS spectra pinpointed the site of pyroglutamate formation as the N-terminus (Figure 2-5B-D). For CBH-1 purified from WT cells, only the pGlu-modified peptide was detected. For CBH-1 purified from the *Δqc-1Δqc-2* mutant, both unmodified N-terminal glutamine and pGlu were detected at an abundance ratio of 9:1, respectively (SI Table 1). The small fraction of pGlu-modified CBH-1 in the Δ*qc-1*Δ*qc-2* mutant may be due to spontaneous cyclization of N-terminal glutamine, a phenomenon previously observed for mammalian proteins (59). Altogether, these data strongly support the hypothesis that QC-1 and QC-2 are essential for the N-terminal pGlu modification of CBH-1 in *N. crassa*.

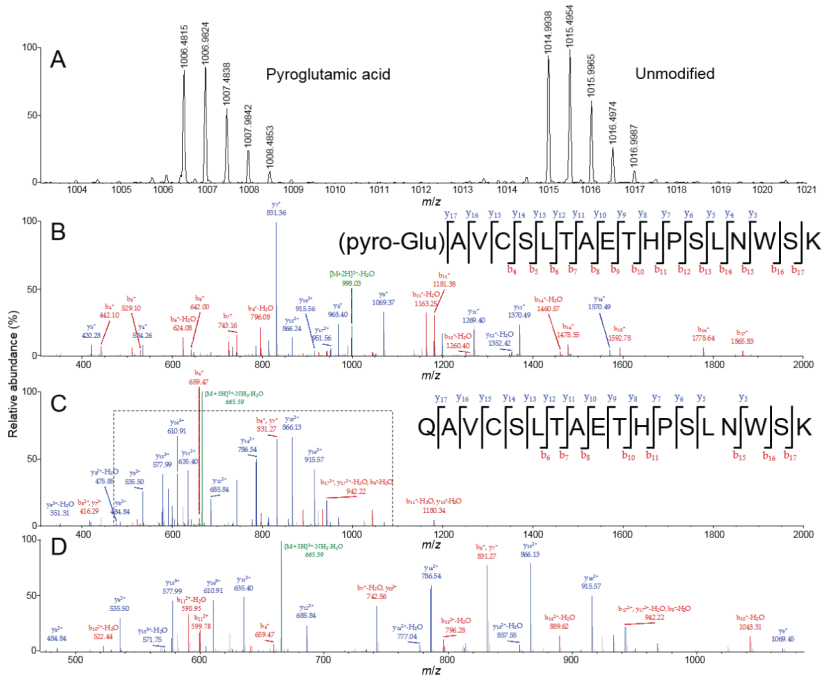


Figure 2-5 LC-MS/MS of *N. crassa* CBH-1. *(A) Full-scan mass spectrum showing detail for isotopically resolved, doubly charged, positive precursor ions occurring at monoisotopic mass-to-charge ratios (m/z) = 1006.4815 and 1014.9938, due to the $[M+2H]^{2+}$ ions of the pGlu modified and unmodified QAVCSLTAETHPSLNWSK peptide. Annotated MS/MS spectra, and peptide sequence maps, resulting from collision-induced dissociation (CID) of precursor peptide ions containing (B) pGlu-modified N-terminus (precursor ion m/z = 1006.4815, $[M+2H]^{2+}$ ion) and (C) unmodified N-terminus (m/z = 676.9987, $[M+3H]^{3+}$ ion). Detail for the region denoted by the dashed line in (C) is shown in (D). MS/MS spectra are annotated using the nomenclature of Roepstorff and Fohlman* (75).

2.4.4. ***pGlu formation occurs in the endoplasmic reticulum (ER)***

In mammalian cells, two QC enzymes catalyze pGlu formation. One QC enzyme has an N-terminal membrane anchor and localizes to Golgi; direct targets of this QC are unknown (78). The second QC is co-secreted with its protein substrates (79); the secreted QC is likely responsible for cyclizing prohormones that mature in secretory granules. We reasoned that *N. crassa* QC-2 resided in the ER due to the

presence of a signal peptide and a KDEL ER-retention sequence (80) (Figure 2-2B). QC-1 also has a signal peptide but no KDEL sequence. The *Δqc-1Δqc-2* strains bearing either *qc-1-sfGFP* or *qc-2-sfGFP* (that fully complemented the CBH-1 phenotype of the *Δqc-1Δqc-2* mutant; Figure 2-4B) were subjected to fluorescence microscopy. QC-1-sfGFP and QC-2-sfGFP fluorescence was observed in structures inside the cell that primarily ringed nuclei, a localization pattern consistent with the ER of filamentous fungi (Figure 2-6A)(81). Differences in the cellular localization of QC-1 or QC-2 in hyphae were not apparent, although strains carrying QC-2-sfGFP had lower GFP signal. To further confirm ER localization of QC-1, co-localization with the ER resident protein SEC61 (82) was performed. A heterokaryotic strain expressing Sec61-mCherry and QC-1-sfGFP showed an identical intracellular localization (Figure 2-6B) indicating that QC-1 localizes to the ER.

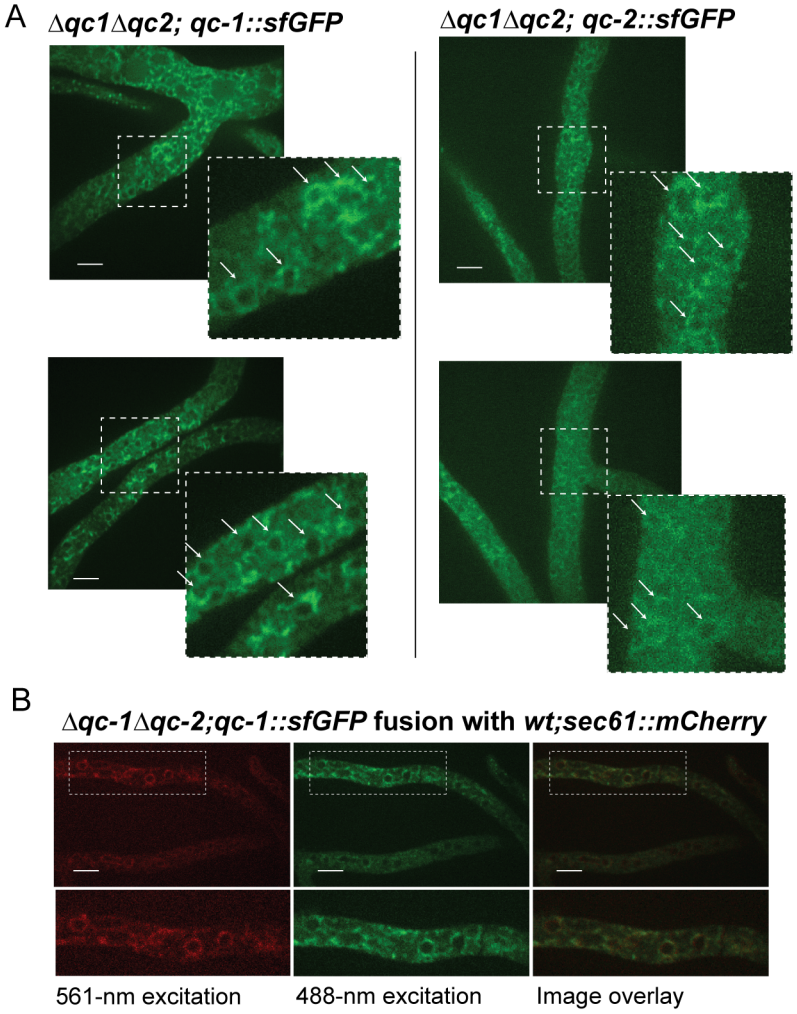


Figure 2-6 Intracellular localization of QC-1 and QC-2. *(A) Fluorescence microscopy of Δqc-1Δqc-2 strains expressing either qc-1-sfGFP or qc-2-sfGFP. (B) Fluorescence microscopy of heterokaryotic cells expressing qc-1-sfGFP and sec61-mCherry* (73)*. Scale bars, 10μm.*

2.4.5. ***pGlu prevents N-terminal truncation of N. crassa endoglucanase GH5-1***

The crystal structure of CBH-1 from a number of filamentous fungi shows an N-terminal pGlu modification (83, 84). The crystal structures of additional GH7 members, including EG1 (Cel7b) (85, 86) and Cel7D (87) have an N-terminal pGlu. A pGlu modification is also observed in the crystal structure of fungal hydrolases outside of the GH7 family, including Cel12A, cellobiose dehydrogenase (Cdh1) and GH10 xylanases (88–91). However, unlike GH7 family enzymes, these enzymes display a pGlu that does not fit into a hydrophobic pocket but rather lies in an extended tail at a distance to the catalytic and substrate binding sites of the protein. This structural aspect of the pGlu modification of *T. reesei* Cel12A is shown in Figure 2-1C, in contrast to GH7 enzymes like CBH-1 (Figure 2-1B).

We therefore investigated endoglucanase II (*gh5-1;* NCU00762), a non-GH family 7 protein that is a highly expressed cellulase in *N. crassa* and has been previously characterized (92). This enzyme has a clear predicted N-terminal glutamine after the signal peptide. To determine whether GH5-1 contained an N-terminal pGlu modification, we constructed strains expressing 6xHis-tagged GH5-1 under the GPD promoter and purified GH5-1 from WT and *Δqc-1Δqc-2* cells. Purified GH5-1 was subsequently subjected to chymotrypsin digestion and LC-MS/MS analysis. As with CBH-1, pGlu-modified GH5-1 was only detected in the WT sample, whereas neither a pGlu-modified nor unmodified GH5-1 N-terminal peptide was detected in the *Δqc-1Δqc-2* sample. Further examination showed that the first 33 amino acids from the N-terminus of GH5-1 were absent in the *Δqc-1Δqc-2* sample (Figure 2-7B)

We reasoned that if GH5-1 from *Δqc-1Δqc-2* cells was N-terminally truncated, it should have a reduced molecular weight as compared to GH5-1 from WT cells. In support of our hypothesis, GH5-1 purified from *Δqc-1Δqc-2* cells showed a band at ~47kD, a lower molecular weight than GH5-1 purified from WT cells (~50kD) (Figure 2-7A). Furthermore, GH5-1 from *Δqc-1Δqc-2* cells ran as a wider-than-expected protein band as compared to GH5-1 from WT cells, suggesting that GH5-1 from *Δqc-1Δqc-2* cells may be subject to truncations at various N-terminal locations.

We hypothesized that GH5-1 could either be truncated extracellularly or by proteolytic processing during passage through the secretory system. To differentiate these possibilities, we subjected whole cell lysates from WT and *Δqc-1Δqc-2* cells to immunoprecipitation and Western blotting with anti-GH5-1 antibody. As shown in Figure 2-7A, no difference in molecular weight was observed between intracellular GH5-1 from WT cells versus the *Δqc-1Δqc-2* cells, indicating that the loss of the N-terminal region of *Δqc-1Δqc-2* GH5-1 occurred post secretion. *N. crassa* has a number of extracellular proteases and peptidases (93), which we hypothesize are responsible for the truncation of unmodified GH5-1.

To test whether truncation of GH5-1 has consequences on enzymatic activity, we subjected purified GH5-1 to a modified filter paper assay. Purified GH5-1 from WT and from *Δqc-1Δqc-2* cells was incubated with 1-cm squares of Whatman cellulose filter paper across a range of temperatures and released sugar-reducing ends were reacted with dinitrosalicylic acid (DNS). GH5-1 from *Δqc-1Δqc-2* cells showed reduced enzymatic activity between 20˚C and 55˚C in comparison to GH5-1

purified from WT cells (Figure 2-7C). At temperatures above 55˚C both enzymes are inactive and show only low levels of endoglucanase activity (Figure 2-7C).

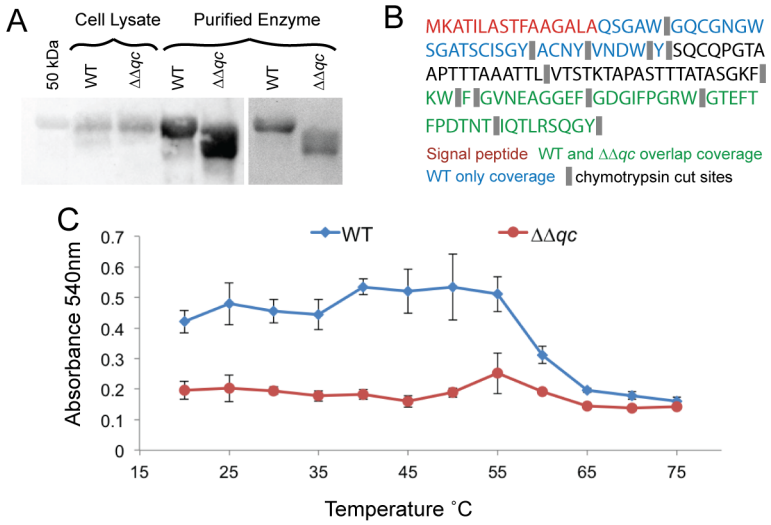


Figure 2-7 Role of pGlu in endoglucanas GH5-1. *(A) Western blot and Coomassie of GH5-1. Right panel shows Coomassie stain of GH5-1 purified from WT and Δqc-1Δqc-2 cells expressing GH5-1-6xHis. Left panel shows Western blot using anti-GH5-1 antibody on cell lysates versus purified GH5-1 from the same strains. (B) Peptide detection by LC-MS/MS of chymotrypsin-digested GH5-1-6xHis expressed in WT versus Δqc-1Δqc-2 cells. Only the first 135 amino acids of the GH5-1 sequence are shown. (C) Endoglucanase activity of purified GH5-1 from WT versus Δqc-1Δqc-2 cells from 20˚C to 75˚C as measured by a filter paper assay* (74)*. Error bars represent standard deviation from quadruplicates.*

2.4.6. ***Pglu can be predicted by N-terminal amino acid***

Human QC cyclizes exposed N-terminal glutamine (Q) or glutamate (E) residues (94). We hypothesized that *N. crassa* proteins with an exposed N-terminal Q or E following signal peptide cleavage would be modified to pGlu when trafficked through the ER. Thus, we sorted predicted proteins destined for the secretory pathway with an N-terminal Q or E (SI Dataset 1); 90 proteins were identified. As expected, CBH-1, GH5-1, CDH-1 and GH7-2 were included in this list. Twelve additional plant cell wall degrading enzymes were identified; many of these enzymes have been detected in the secretome of *N. crassa* (95–97). The two largest remaining groups are proteins involved in modulation of the fungal cell wall (98–100) and secreted proteases. The remaining proteins have a range of function including oxidoreductases, nucleotidases, phosphatases, the ER resident protein IRE-1 and several vacuolar proteins. Finally, there are 47 proteins with no known functions, 19 of which show conservation among ascomycete fungi outside of Sordariomycetes.

2.5. **Discussion**

Secreted glutaminyl cyclases in mammals play a role in stability, function and aggregation of secreted peptides and proteins (58, 59, 62, 63). pGlu modification is required for certain peptides for resistance to peptidases and to remain receptor-active (59, 101). Here, we demonstrated that pGlu modification of secreted cellulase enzymes in *N. crassa* is catalyzed by two QCs that reside in the ER. Deletion of *N. crassa qc-1* and *qc-2* resulted in a significant reduction in pGlu-modified CBH-1 and GH5-1. In addition, CBH-1 and GH5-1 lacking the pGlu modification showed a reduction in enzyme activity and unmodified GH5-1 was subject to N-terminal proteolytic degradation. These data support a role of pGlu in protecting the N-termini of secreted proteins from degradation by endogenous and exogenous exopeptidases.

A second QC in mammals is retained within the Golgi complex (isoQC) (78, 102). Differential cellular distribution of QC and isoQC may indicate a preference for different substrates in the secretory pathway and thus distinct physiological roles. This has been demonstrated in monocyte chemo-attractant proteins (MCPs) versus thyrotropin-releasing hormone (TRH) where only one QC is involved in the pGlu formation for each protein (102, 103). In *N. crassa,* both QC-1 and QC-2 localized primarily to the ER, where proteins with a liberated N-terminal Q or E would be cyclized to pGlu. It is unclear whether QC-1 or QC-2 may preferentially catalyze pGlu on particular proteins. Future studies on elucidating further and potentially specific targets of these two QC proteins in *N. crassa* will be informative.

TeCBH1 and TeEG1 heterologously expressed in *S. cerevisiae* show reduced thermal stability and presumably lack the N-terminal pGlu modification (22, 64). In hindsight, this result is surprising because a QC homolog is present in the *S. cerevisiae* genome (which is uncharacterized, but a GFP-tagged version is reported to localize to the ER; YFR018C)(104) (SI Fig. 1). However, to enhance expression in *S. cerevisiae*, an engineered α factor APPS4 pre-pro leader sequence (105) was appended to the N-terminus of the *TeCBH1* and *TeEG1* constructs. The leader sequence APPS4 contains a secondary KEX2 site after the signal peptide cleavage site, causing the N-terminal Q to only be revealed at the Golgi where KEX2 cleavage occurs. If *S. cerevisiae* QC is ER localized like *N. crassa* QCs, pGlu formation would not occur when using APPS4 as a leader sequence, resulting in TeCBH1 and TeEG1 proteins that lack this modification and thus thermally less stable.

Fungal systems may prove useful for the study and production of pGlu modified proteins. Recent studies have indicated that pGlu modified protein production is a difficult process, where complex expression systems or post expression treatment with purified QC is required (106, 107). As shown in this study, simple procedures can be used to tag, express and purify secreted proteins from WT and from Δ*qc-1*Δ*qc-2* strains. Compared to other expression systems, filamentous fungi have the advantage of being eukaryotic, with high protein production strains (108). Appending pGlu to non-modified proteins may be a useful tool for heterologous expression of proteins in filamentous fungi for biotechnological applications. Where the N-termini are not involved in essential functions, pGlu addition could maintain protein stability especially in systems where exopeptidases are ubiquitous.

We identified ~90 proteins from *N. crassa* predicted to be pGlu modified. The majority of these proteins are predicted to be secreted and include plant cell wall degrading enzymes, proteases, fungal cell wall proteins and putative small secreted peptides of unknown function. Plant pathogenic fungi secrete a diverse set of small, secreted effector proteins crucial for pathogenesis (109). Since many pathogenic fungi also have genes encoding QCs in their genome (Fig. S1), it is possible that the pGlu plays a role in the function/stability of these effector proteins. We also identified a few intracellular proteins with a predicted pGlu modification, including the ER protein Ire1p, which is important for triggering the unfolded protein response (110). The predicted pGlu site is well conserved among *IRE1* homologs in filamentous ascomycete fungi, but is not conserved in *S. cerevisiae* Ire1p. The N-terminal domain of Ire1p has been shown to be important for binding to hydrophobic regions of unfolded proteins (110), and pGlu hydrophobicity in IRE-1 in filamentous fungi could potentially play a role in this process. Future experiments to dissect the role of pGlu modification in secreted and intracellular proteins will address these questions.

Chapter 3. **Transcriptional profiling of *N. crassa* exposed to various carbon, nitrogen, sulfur and phosphate sources**

3.1. **Introduction to profiling**

The initial goal of this project was to understand the overarching principles that drive fungal nutrient response and acquisition. How nutrients are detected, elicit cellular responses, transport and catabolism are questions that researchers have worked on for a long time. A better understanding of nutrient acquisition will allow researchers to intelligently engineer filamentous fungi for many purposes.

In order to reveal the full story of nutrient uptake, we first wanted to understand how fungi respond to the presence of different nutrients in their environment. We used *N. crassa* as a model for filamentous fungi. We investigated how *N. crassa* responds to different nutrients by assessing gene expression patterns using RNAseq. In collaboration with the Joint Genome Institute we examined *N. crassa's* response to a large number of different carbon, nitrogen, phosphate and sulfur sources (see section 3.2.1).

In order for our conditions to be comparable, we tried to find a high throughput and consistent method to access transcription of cells exposed to various nutrient sources. We settled on a protocol where we grew small volumes of mycelia on sucrose minimal media overnight, washed the cells of all nutrients, and transferred cells into new media for a 4-hour induction phase (Figure 3-1). Further details are described in the materials and methods section of this chapter.

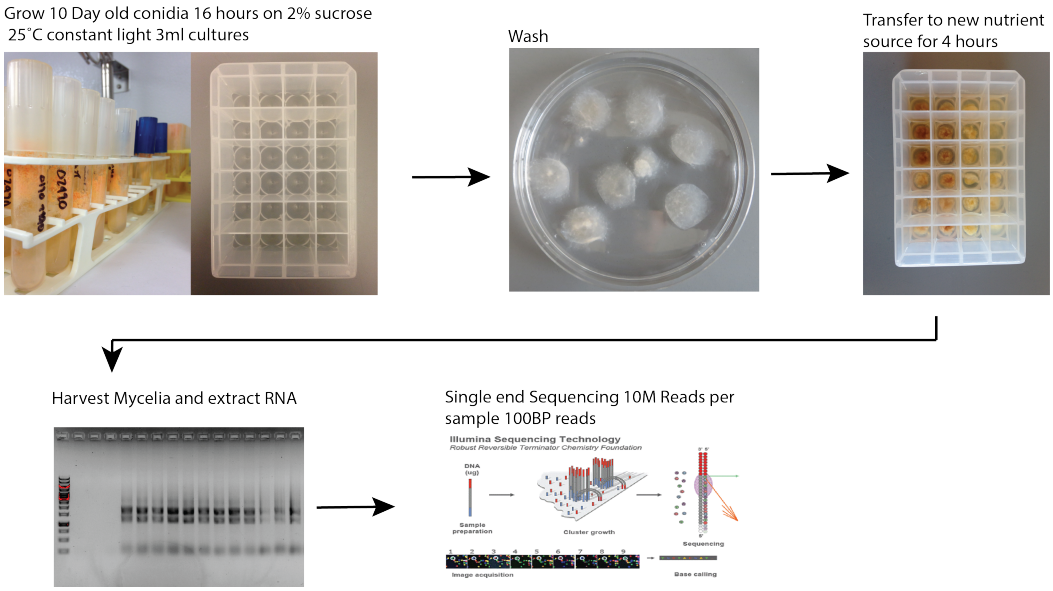


Figure 3-1 Schematic for growth, induction and RNAseq analyses

After profiling *N. crassa* WT strain (FGSC 2489) in ~50 conditions, we evaluated the data to generate new hypotheses regarding genes that might be involved in nutrient sensing, acquisition, growth and gene regulation. Comparisons between different conditions helped us understand which nutrients elicit strong or weak responses. This approach allowed us to distinguish metabolic priorities and preferences, especially regarding carbon acquisition. We turned to the expression profiles of more specific sets of genes to find candidate transcription factors, transporters, and other genes that we hypothesized were involved in catabolism of specific nutrients. This chapter focuses on the general conclusions we drew from our WT dataset, as well as the specific genes of interest we found using the transcriptome profiling technique. With the help of Sara Calhoun at the JGI, we were also able to generate a gene co-expression network, grouping genes that showed similar expression patterns across the conditions we tested.

3.2. Materials and Methods

3.2.1. *Growth and RNA extraction*

General Growth and RNA extraction

N. crassa WT (FGSC 2489) and gene deletion strains were obtained from the Fungal Genetics Stock Center (FGSC; [www.fgsc.net]). Conidia obtained from 10-day-old pre-grown cultures were used to inoculate 3ml of 1 x Vogel's salts plus 2% (w/v) sucrose at $1x10^6$ cells/mL in 24 well Whatman Uniplates. The bottoms of the wells for each Uniplate were initially scratched with a sharp needle to allow adherence and formation of mycelial mat. After 16 hours of growth, mycelia were washed three times in 1 x Vogel's salts (111) without added carbon and subsequently transferred to 1 x Vogel's salts with a new carbon, nitrogen, phosphate or sulfur source (described below). After 4 hours of induction, mycelia were

harvested over Whatman #1 filter paper and subsequently flash frozen in liquid nitrogen for storage at -80˚C. Total RNA was isolated with TRIzol reagent (Invitrogen), treated with TURBO DNase (Thermo Fisher), and purified using Qiagen RNAeasy Purification Kit. RNA was tested for quality using agarose gel electrophoresis.

Carbon conditions

Our growth conditions consisted of a number of carbon sources. Mono and disaccharides were used under inducing conditions at 2mM (see section 3.3.1), while complex polysaccharides and plant biomass were used at 1% (w/v). Plant biomass were obtained by grinding plant material from those sources to ~.08mm size. Carbon conditions and sources from which they were obtained (including CAS# if available) are listed below:

Condition	Source and Specifications	CAS Number
lactose	Sigma β lactose >99%	5989-81-1
fructose	Research organics	0609-06-03
xylose	Acros Organics 225990050	58-86-6
Mannose	Acros D (+) 99+%	3458-38-4
maltose	Sigma D (+) min 98% <0.3% glucose <1.0% maltotriose	6363-53-7
arabinose	Sigma L (+) min 99% a3256	5328-37-0
cellobiose	Fluka D (+) >99%	
galactose	Acros D (+) 99+%	59-23-4
rhamnose	TCI R0013	10030-85-0
galacturonic acid	Fluka 48280	
glucuronic acid	Sigma 98%	12/3/56
trehalose	Acros D 99%	6138-23-4
mannobiose	Megazyme	O-MBI
sorbitol	Sigma	50-70-4
sorbose	Calbiochem L(-) 99.6%	CAS 87-79-6
glycerol	Fisher	56-81-5
ribose	TCI R0025	50-69-1
mannitol	Fisher	69-65-8
sodium acetate	Fisher 5210	
fucose	Sigma	2438-80-4
inulin	Sigma	9005-80-5
Avicel	Fluka	9004-34-6
xylan	Sigma from beechwood	9014-63-5
xyloglucan	Megazyme (from tamarind)	
galactomannan	Megazyme (carob low viscosity)	
glucomannan	Konjac Foods	
mixed linkage glucan	Megazyme (barley	

pectin	Sigma from orange peel	9000-69-5
pectin esterified	Sigma	9046-40-6
polygalacturonic acid	Sigma	25990-10-7
rhamnogalacturonan	Megazyme (potato)	
arabinan	Megazyme (sugar beet)	
galactan	Megazyme (lupin)	
amylopectin	Sigma from corn A-7780	
amylose	Sigma from corn A-7043	
mannan	Megazyme (ivory nut)	
xyloglucan	Megazyme	
Miscanthus	in house	
Switchgrass	in house	
Corn Stover	in house	
Energy Cane	in house	
Citrus peel	in house	
Wing Nut	in house	
Poplar	in house	
Lobloly Pine	in house	
Willow	in house	

Nitrogen, phosphorous and sulfur conditions

For nitrogen starvation condition, Vogel's salts (111) lacking ammonium nitrate was used. This stock was used to make medias with individual amino acids as the sole nitrogen source. Amino acid concentrations were as follows: glutamine 50mM, glutamate 50mM, proline 43mM, alanine 112mM, glycine 200mM, arginine 344mM, BSA 1% w/v, NaNO3 50mM.

To make sulfur free media, magnesium sulfate was replaced with magnesium chloride in Vogel's salts. For low sulfur media, sulfur free media was adjusted to 0.25mM concentration of methionine. 5mM methionine was used for high sulfur media (112).

To make phosphate free media, low phosphate media and high phosphate media, we a base media that contained: 2% w/v sucrose, 5.0g/L ammonium tartrate, 1g/L ammonium nitrate, 0.5g/L magnesium sulfate heptahydrate, 0.1g/L calcium chloride, 0.1g/L sodium chloride and additional supplements of trace elements and biotin. For phosphate free media, 2.5g/L potassium chloride was added to this base media. For low phosphate media, 50mg/L potassium phosphate was added. For High phosphate or regular Fries media, 1g/L potassium phosphate was added (113).

3.2.2. *RNAseq*

Library preparation and sequencing

Stranded cDNA libraries were generated using the Illumina Truseq Stranded RNA LT kit. mRNA was purified from 1 μg of total RNA using magnetic beads containing poly-dT oligos. mRNA was fragmented and reversed transcribed using random hexamers and SSII (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 10 cycles of PCR.

The prepared library was then quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified library was then multiplexed with other libraries, and the pool of libraries was then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and Illumina's cBot instrument to generate a clustered flowcell for sequencing. Sequencing of the flowcell was performed on the Illumina HiSeq2000 sequencer using a TruSeq SBS sequencing kit, v3, following a 2x100 indexed run recipe.

Analysis

Filtered reads were mapped against *N. crassa* OR74A genome (v12) using Tophat 2.0.4 (114). Transcript abundance was estimated with Cufflinks 2.0.2 (115) in fragments per kilobase of transcript per million mapped reads (FPKMS) using upper quartile normalization and mapping against reference isoforms from the Broad Institute. Tophat mapped reads were additionally counted by HTSeq 0.6.0 (116) to obtain raw counts. Differential expression analysis was preformed on raw counts with DEseq2 version 3.3 (117) using data from biological triplicates.

3.2.3. *Post RNAseq analyses*

Hierarchical clustering

FPKMs were hierarchically clustered with Cluster 3.0 software (118). FPKMS were log transformed and normalized prior to clustering. Clusters were generated using uncentered correlation as the similarity metric, and average linkage as the clustering method. Clustering heatmaps were viewed and exported from Java Treeview (119).

MFS transporter tree construction

A list of all predicted transporters in *N. crassa* was obtained in collaboration with the JGI. A pipeline for transporter prediction using data from the transporter classification database (TCDB) (http://www.tcb.org/), was used to generate this list. Major Facilitator Superfamily transporters were selected from this list and protein sequences for all MFS transporters obtained from Broad v12 genome protein sequence .fasta file. Sequences were aligned using MAFFT version 7 (68). The alignment was used to construct a maximum likelihood phylogeny using RAxML (69). FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used for visualization.

3.2.4. *WGCNA*

The gene co-expression network was calculated across expression profiles for the wild-type strain grown in 72 different conditions using the R package WGCNA (Langfelder and Horvath, 2008). After filtering genes out due to low expression (<10 FKPM) across >95% of all conditions, 6,742 genes were used in correlation analysis. The correlations were scaled using soft power of 9, assuming a scale-free network. Hierarchical clustering was applied to identify co-expressed gene modules with a minimum module size of 30 genes. The network with a force-directed layout was visualized using Cytoscape version 3.3 (120). The scaled correlations between gene pairs above 0.15 are represented as edges in the network.

FUNCAT

A custom catalogue (35) based on MPS Functional Catalogue Database (FuncatDB) with expanded categories for cell wall degradation realted genes was used for enrichment analysis. Functional enrichment for the gene modules was performed using functional assignments in FunCatDB (latest version) (121). The enrichment was calculated by the hypergeometric test and adjusted for multiple hypothesis testing using the Bonferroni correction. CAZy gene annotations were obtained from the JGI MycoCosm portal, Neucr2 (122).

3.3. Carbon profiling

We challenged *N. crassa* with a large selection of various carbon sources. These consisted of simple sugars, complex polysaccharides and full plant biomass. We focused our attention on expression patterns of genes encoding PCWDEs. These transcriptome profiles gave us much insight into how PCWDEs are induced and regulated. Different sugars tended to induce different sets of PCWDEs, indicating that *N. crassa* utilizes many different sugars as signaling molecules, rather than just a few. At the same time, some sugars induce the same sets of PCWDEs, indicative that they signal the same regulatory network. Expression changes of PCWDEs induced by sugars reflect the plant cell wall polysaccharides that these sugars are derived from. Comparison of PCWDE expression profiles between polysaccharides and their constituent sugars lead us to believe that many of polysaccharides sources are impure and contain other plant cell wall components. These comparisons also inform us that cellulase production is a higher priority process than hemicellulase or pectin production in *N. crassa.* Finally we observe much overlap in the PCWDEs induced by different carbon sources. Even at the sugar level, monomers from different cell wall components can induce expression of the same PCWDE genes. This observation is indicative that activation of PCWDEs is not regulated in independent gene sets, rather PCWDEs can likely be induced by multiple regulators.

3.3.1. ***N. crassa has unique responses to a broad range of sugars, but responds strongly to select few.***

The carbon sources that we chose to profile *N. crassa* with were divided into three categories: Simple sugar monosaccharides and disaccharides, complex

polysaccharides, and full plant biomass. For the sugar monosaccharides and disaccharides, we chose a diverse set of molecules present as building blocks of the plant cell wall (*see materials and methods*). We hypothesized that sugar monosaccharides and disaccharides or downstream products from the catabolism of these sugars are the molecules that serve as signals for PCWDE expression. While large polysaccharides cannot enter the cell, sugar monosaccharides and disaccharides can be specifically transported into the cell and thus a likely signal for the cell to initiate production of PCWDEs. Previous literature has shown that certain ascomycete fungi respond to cellobiose and xylose to induce expression of cellulases and hemicellulases. *Aspergillus* expresses *cbhA* and *cbhB* when induced by a low concentration of xylose (123) and work from our lab has shown the same is true with cellobiose and *cbh-1* in *N. crassa* (43, 124). Unknown is whether fungi use many additional sugars as signals or only these few unique sugars including cellobiose and xylose to mount production of PCWDEs.

We used cellobiose as a proxy to determine the concentration of sugars to use in order to induce PCWDE transcription without causing carbon catabolite repression (CCR). Elizabeth Znameroski, a previous graduate student in the lab, demonstrated that a narrow concentration range of cellobiose could induce transcription of the *cbh-1*. *Cbh-1* was not highly transcribed at concentrations 5.8mM and above or 100uM and below (124). Presumably, concentrations above 5.8mM induced CCR and thus repress expression of cellulases, while below 100uM was under the threshold for detection by the cell. Elizabeth determined that 1mM was a good concentration for *cbh-1* induction in 100ml shaking flask cultures. We ran similar experiments using cellobiose and assessed *cbh-1* induction to determine the best concentration for induction of PCWDEs in 3ml cultures. We tested several concentrations above and below 1mM and determined that 2mM most consistently induced the expression of *cbh-1* approximately 200-fold as compared to no carbon conditions (Figure 3-2). From these data, we chose 2mM concentration for all the sugar monosaccharides and disaccharides tested.

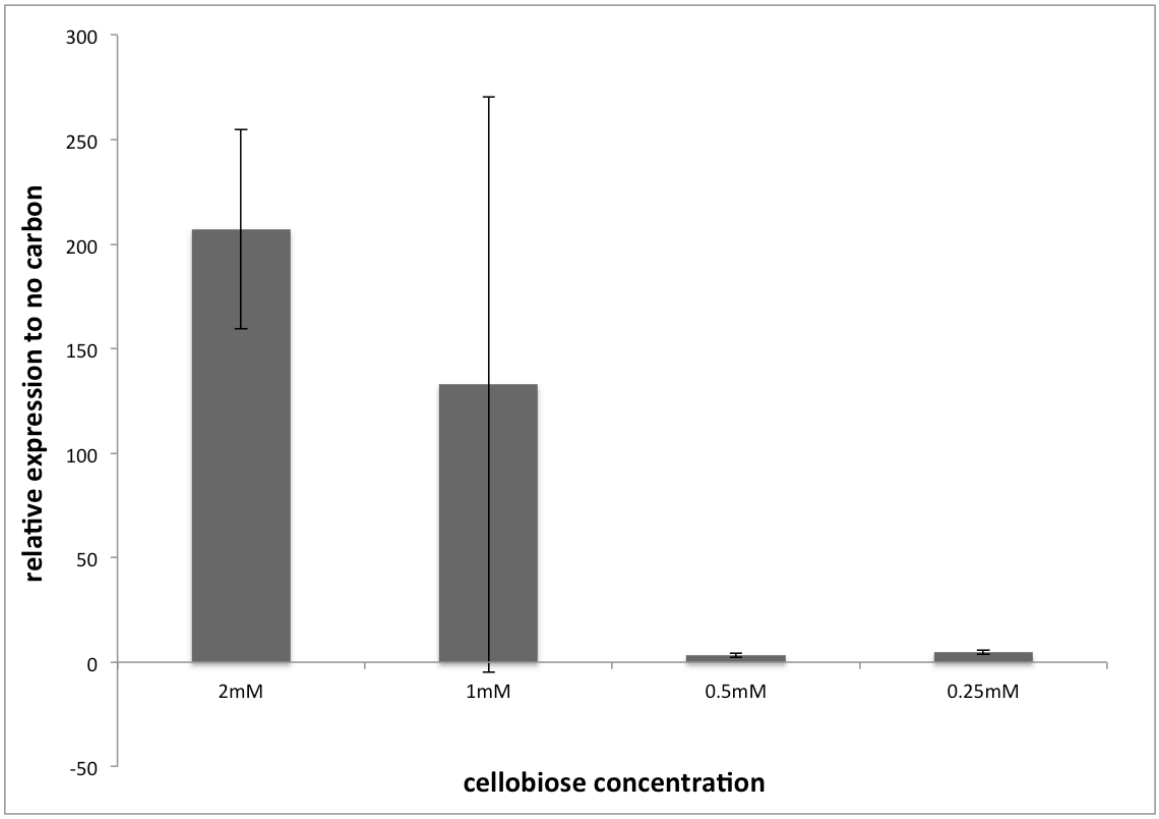


Figure 3-2 Induction of *cbh-1* by varying concentrations of cellobiose *qRT-PCR results of cbh-1 transcript abundance in varying concentrations of cellobiose relative to no carbon condition in triplicate. Error bars represent standard deviation from biological triplicate.*

To compare responses to different sugars, we conducted RNAseq cells induced by each of our sugar conditions and used hierarchical clustering to examine expression of carbohydrate active enzymes (CAzymes) in order to determine whether unique sugars induce unique transcriptional responses. We determined that *N. crassa* responds more strongly to certain sugars such as cellobiose, xylose and trehalose. The fungus is additionally able to detect other sugars and responds accordingly with up-regulation of specific sets of genes encoding CAzymes. Two exceptions were lactose and sorbitol. These two sugars induced expression profiles almost identical to *N. crassa* on no carbon, thus indicating that the cell either does not detect or does not mount a response to these specific sugars. These data also indicated that carbon catabolite repression is likely not occurring with any of these sugars at 2mM concentration. Only under 2% sucrose conditions are the majority of genes encoding CAZymes repressed (Figure 3-3).

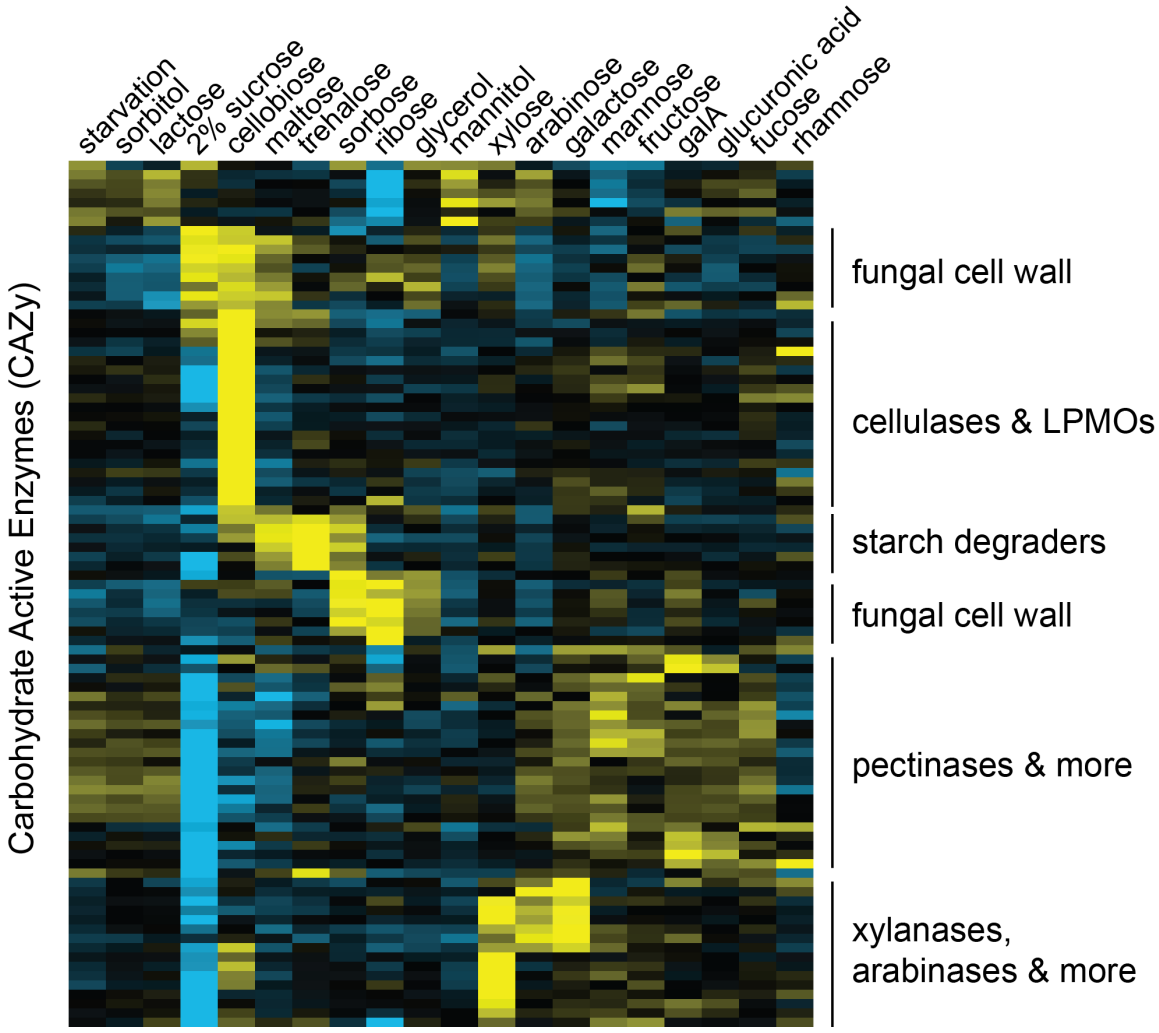


Figure 3-3 Hierarchical clustering of genes encoding CAZymes across 2mM mono and disaccharides RNAseq datasets *Mean normalized counts from biological triplicates were used for clustering analyses. Blue represents down regulation and yellow up regulation.*

3.3.2. ***Complex polysaccharides***

We chose to profile the transcriptional response of *N. crassa* on a wide range of complex polysaccharides, including representatives of the major polysaccharides present in the plant cell wall. This included components of starch, cellulose, hemicellulose and pectin. We found that different polysaccharides induced different sets of genes encoding PCWDEs, similar to what we had observed in 2mM sugars. Like the sugars, certain polysaccharides elicited a much stronger response than others. There was also significant overlap in the CAZymes up regulated across different polysaccharides. Most surprising was the induction of a number of genes encoding cellulases induced by exposure to pectin and homogalacturonan (HG). Unlike the sugar data, polysaccharides tended to induce a large number of CAZymes as expected by the greater complexity of the substrate (Figure 3-4).

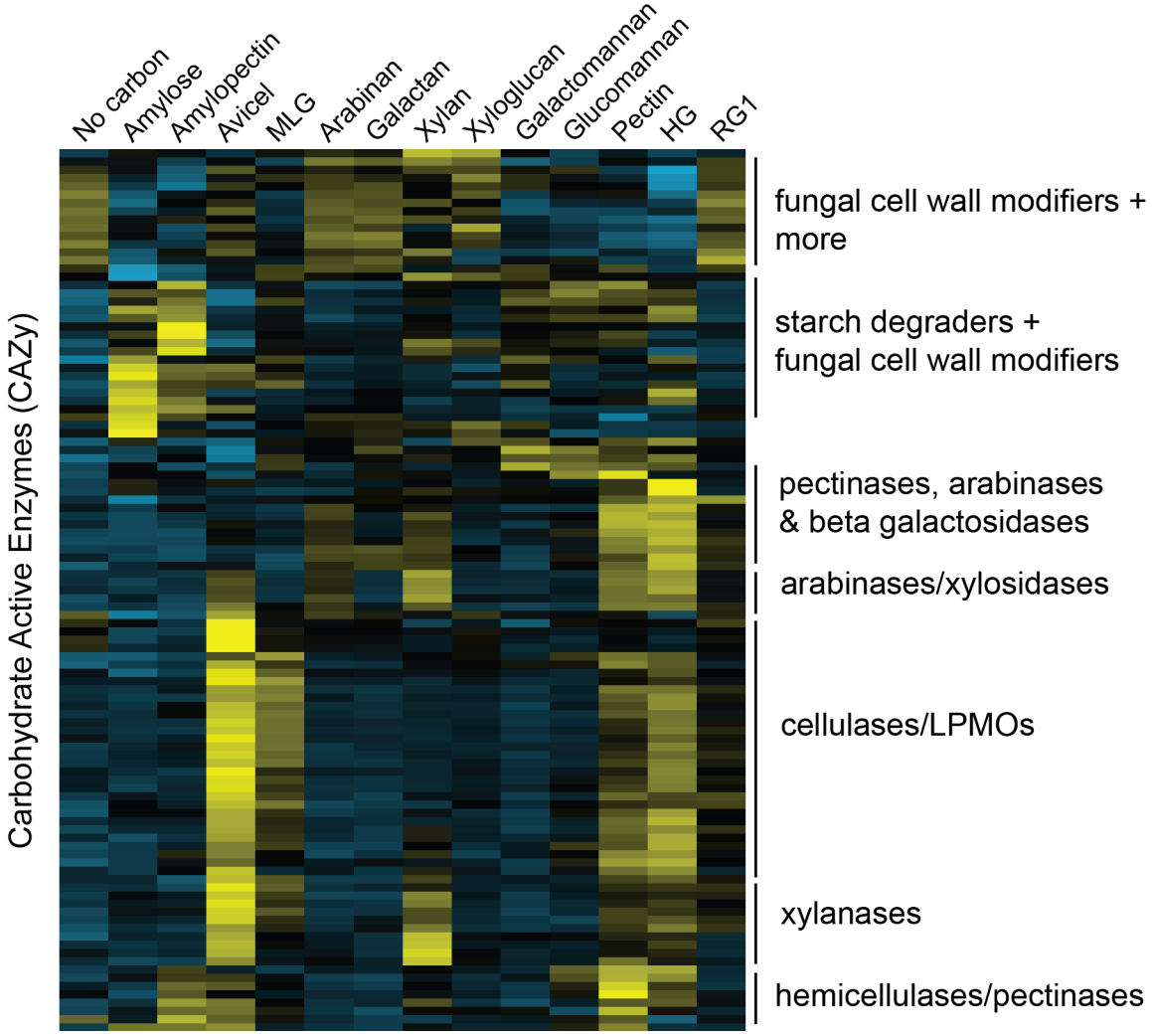


Figure 3-4 Hierarchical clustering of genes encoding CAZymes under induction by complex polysaccharides *Mean normalized counts from biological triplicates were used for clustering analyses. Blue represents down regulation and yellow up regulation.*

We found that the components of starch, amylose and amylopectin, both elicit expression of genes encoding starch-degrading enzymes, but amylopectin is a stronger inducer. Under amylopectin, the α-amylases (NCU09805, NCU05873), trehalase (NCU00943) and glucoamylase (NCU01517) are all up regulated, while under amylose conditions only the amylase *amy-1* (NCU05873) is up regulated. Interestingly, a fair number of fungal cell wall modifying enzymes are also induced under these two starch conditions. It is possible that the properties of starchy environments force the mycelia to grow differently than it would under other conditions.

We found that exposure to Avicel induces more genes than cellulases and additional PCWDEs. In addition to genes encoding cellulases, Avicel induces a handful of hemicellulases and pectinases. This data indicates that either Avicel is contaminated with other polysaccharides, or that transcription factors induced by

cellulose also positively regulate these other PCWDEs, or both. There is additionally a cluster of endoglucanases and β-glucosidases (NCU03328, NCU05121, NCU04854, and NCU04952) that is very specific for Avicel and does not appear to be up regulated in any of our other polysaccharide conditions, even though other cellulases are up regulated under these conditions.

N. crassa exposed to arabinan and galactan has expression profiles very similar to starvation. Under starvation, there is only a single cluster of genes that is up regulated compared to the other conditions. This set contains a mix of genes encoding CAZymes that consist mostly of fungal cell wall modifiers. This cluster is also up regulated under arabinan and galactan indicating similarities between starvation and these two conditions. Furthermore, all the CAZymes present in the plot are down regulated in arabinan and galactan as observed with no carbon except for a few exceptions. Under arabinan, arabinases (NCU05965, NCU09170, NCU09775, NCU09924, NCU02343), several β-galactosidases (NCU00642, NCU04623) and pectinases (NCU02369, NCU06961) show minor up regulation, but not at levels observed when *N. crassa* is exposed to pectin and HG. On galactan, the same two pectinases described above are up regulated, but surprisingly, none of the β-galactosidases are up regulated. Perhaps *N. crassa* uses arabinan/arabinose to specifically signal for production of galactosidases.

Exposure of *N. crassa* to pectin and HG results in a surprising CAZyme profile. Exposure to pectin not only induces pectinases, but also induces a large number of hemicellulases as well as cellulases. As seen in Figure 3-4 pectin and HG have the largest set of up regulated genes. These data indicate that either pectin and HG are highly contaminated with other polysaccharides or that the TFs induced by pectin regulate a depolymerization of other polysaccharides or both.

3.3.3. ***Constituent sugars of polysaccharides are enough to induce the majority of PCWDEs induced by the polysaccharide itself***

To determine whether the responses induced by complex polysaccharides are similar to the responses elicited by their constituent sugars, we examined the responses to Avicel, xylan, pectin, amylose and amylopectin and compared them to their constituent sugars and polysaccharides. Direct comparisons were made between Avicel to cellobiose, xylan to xylose and arabinose, pectin to HG, RG1, GalA, galactose and rhamnose, and amylose/amylopectin to maltose and trehalose. We compared differential expression of PCWDEs between no carbon versus a polysaccharide with the same data from no carbon versus a constituent sugar. Along with PCWDEs, we included important intracellular enzymes that catalyze the metabolism of these constituent sugars. We hypothesized that there should be induction of similar genes between polysaccharide and its constituent sugars. We found that PCWDE induction by polysaccharides is indeed a sum of its parts: constituent sugars combine to induce PCWDEs at levels observed under their parent polysaccharides.

Cellulose (Avicel) and cellobiose induction of PCWDEs

A comparison between the Avicel and cellobiose regulons showed that these two carbon sources induce a similar set of cellulases (Figure 3-5). When we

examine the list of cellulases more closely, we see that their gene expression profiles do not match up completely. Certain cellulase genes are more highly induced on Avicel than cellobiose and visa versa. Genes more highly induced on Avicel might be responding to other signals besides cellobiose present in Avicel.

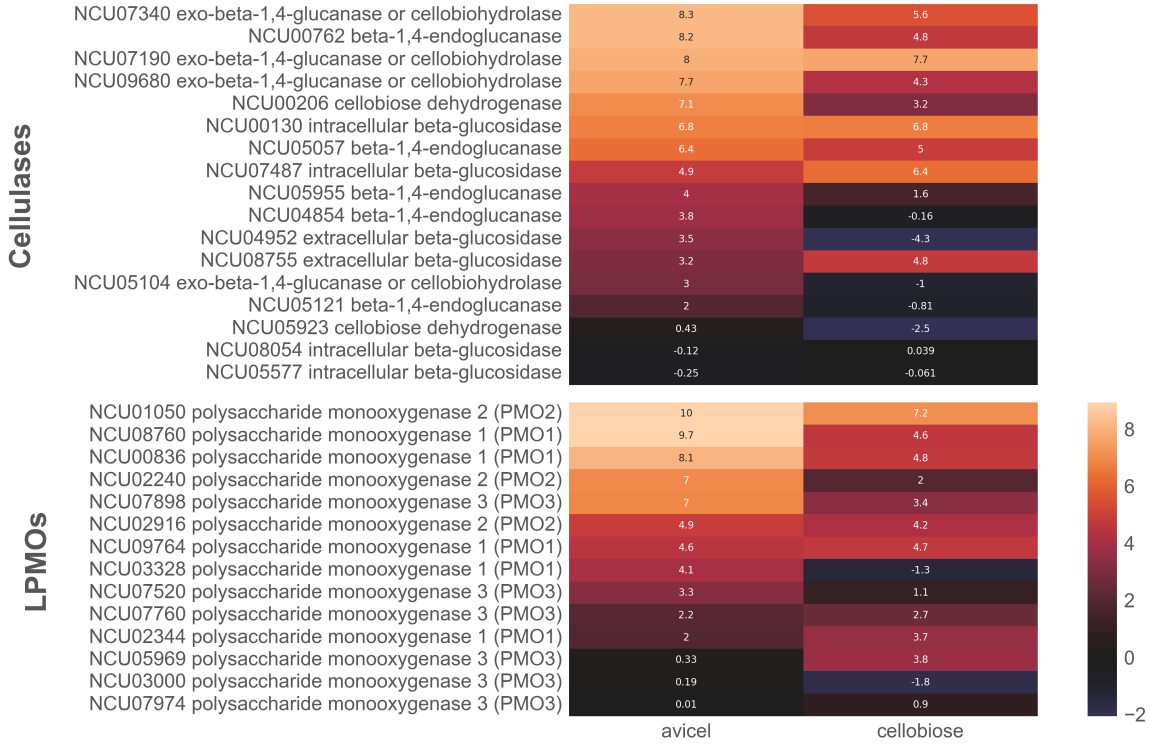


Figure 3-5 Comparison of cellulase gene expression under 1% Avicel versus 2mM cellobiose. *Heat map shows differential expression as measured by log_2 fold change between no carbon vs Avicel (left column) and no carbon vs cellobiose (right column). Log_2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar. Genes are ordered by descending value of log_2 fold change between no carbon and Avicel.*

We also compared expression of no cellulase PCWDEs under Avicel and cellobiose, and observed that both carbon sources induce depolymerization of non-cellulose polymers, but to a greater extent under Avicel induction (Figure 3-6). We observed high induction of many of these hemicellulases and pectinases under both conditions. It was surprising for us to observe high induction under cellobiose conditions, but not as surprising under Avicel conditions. Due to its extraction from plant matter, Avicel is contaminated with other plant cell wall polymers (43). Cellobiose, may also be impure but impurities are below 1%, so any impurities would be in the μm range in the final media, which we have shown is unlikely to induce expression of PCWDEs (Figure 3-2). These data therefore indicates that cellobiose truly induces expression of a whole suite of genes encoding hemicellulases and pectinases. This observation suggests that *N. crassa* regulates expression of PCWDEs genes in a complex and integrated manner; where multiple signals combine to fully induce expression of any given PCWDE.

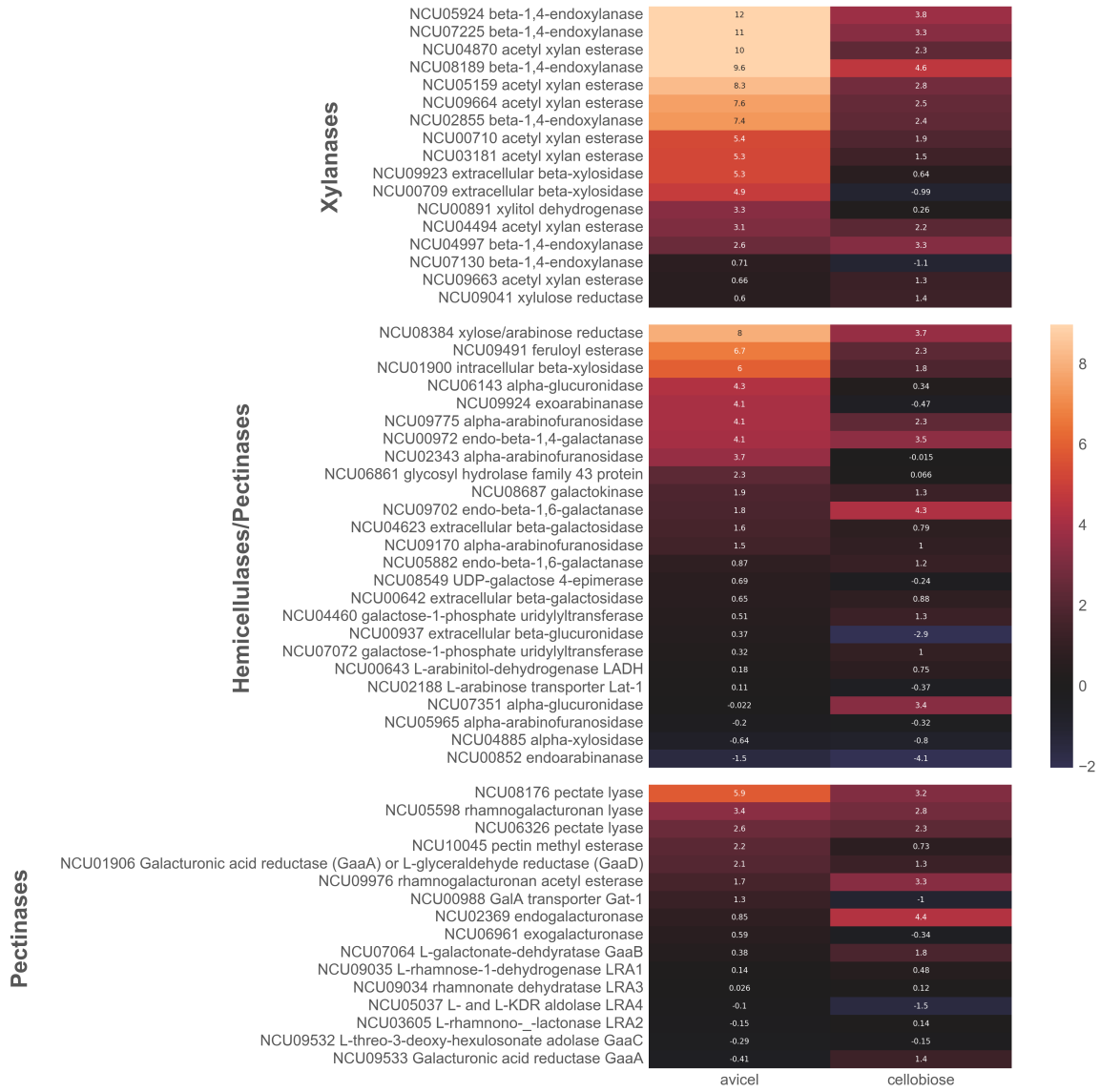


Figure 3-6 Comparison of hemicellulase and pectinase gene expression under 1% Avicel vs 2mM cellobiose. *Heat map shows differential expression as measured by log_2 fold change between no carbon vs Avicel (left column) and no carbon vs cellobiose (right column). Log_2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar. Genes are ordered by descending value of log_2 fold change between no carbon and Avicel.*

Our comparison of genes encoding PCWDE induction by Avicel versus cellobiose informs us that cellulose is a very strong inducer of a wide range of PCWDEs and not just cellulase genes. Cellobiose signaling explains the majority of cellulase gene expression, as we observe a very similar set of PCWDEs induced on Avicel and cellobiose. These data are also in accordance with previous findings that describe the two major cellulase regulators CLR-1 and CLR-2, which become activated in the presence of cellobiose and induce expression and secretion of the

majority of cellulases (27, 28, 43). However, we also observed that not all cellulases are induced to the extent that they are on Avicel, indicating that they may be more molecules that signal transcription of cellulase genes and that may act through alternate transcription factors.

Xylan and its constituent sugars

We conducted similar analyses with xylan, xylose and arabinose as we did with Avicel and cellobiose. Xylose and arabinose make up the majority of sugars present in xylan. We observed that the combined induction by these two constituent sugars induces the majority of xylanase gene expression observed by exposure to xylan (Figure 3-7).

Xylose is to be the major inducer of xylanases and 2mM xylose induces most xylanases to approximately the same levels as xylan. Arabinose plays a minor signal for xylanase induction, but is required to achieve full induction of certain xylanase genes (NCU00709, NCU09923, NCU05924). Other xylanase genes (NCU09663, NCU03181, NCU07130) are not induced by xylan, xylose or arabinose. These genes are likely induced by other signaling molecules. Indeed Avicel and cellobiose induce expression of NCU03181, an acetyl xylan esterase.

Xylose and arabinose also play a minor role in pectinase gene induction. Both xylose and arabinose are present in the side chains of pectic polysaccharides. We observe that a small number of pectinases are induced by arabinose and xylose as seen in the two lower data sets of Figure 3-7. Many of these pectinase genes are also induced under xylan, but similar to Avicel, the purity of xylan from other cell wall polymers is unreliable. Xylan could very well be contaminated with pectinaceous polysaccharides.

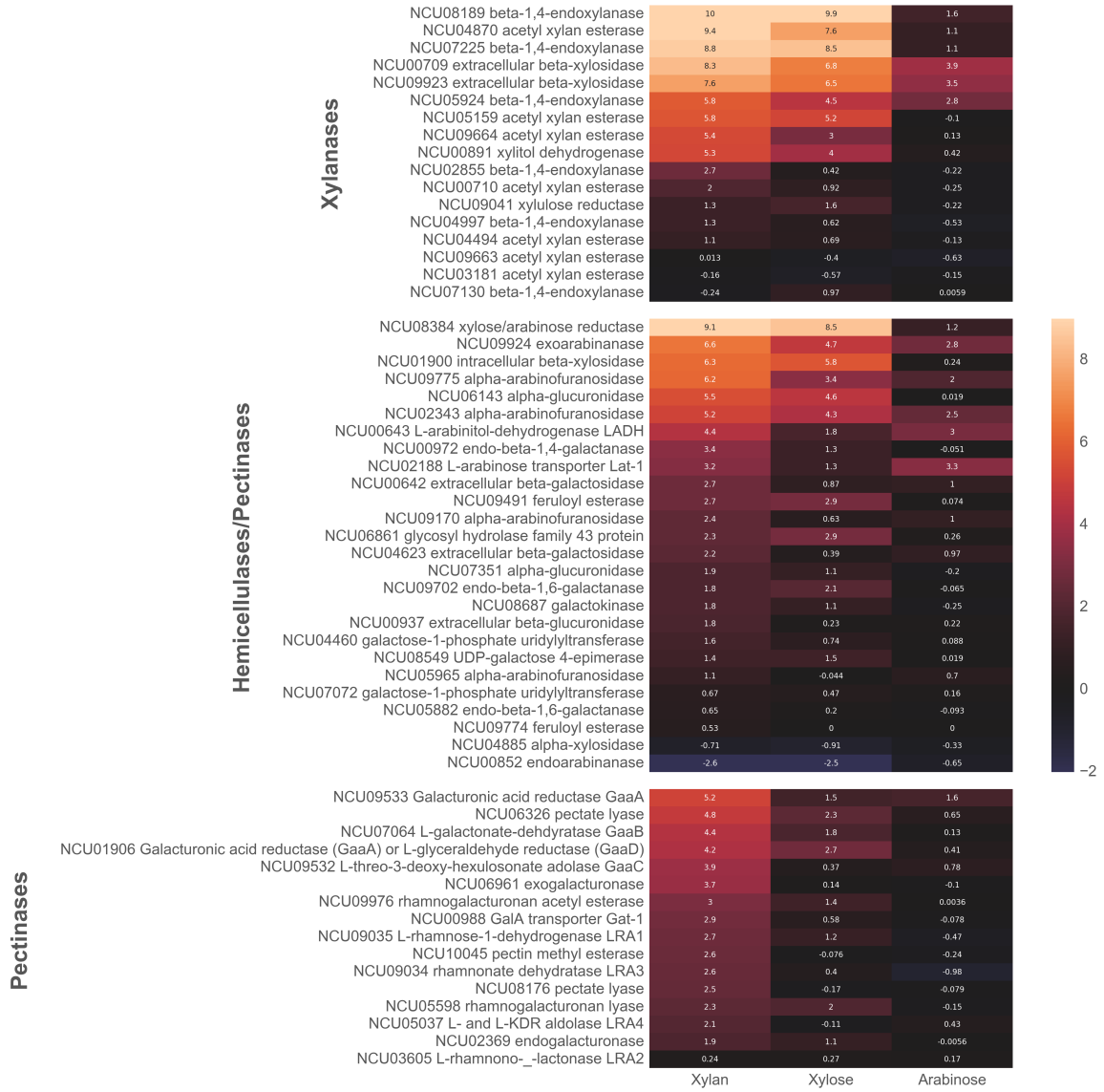


Figure 3-7 Comparison of hemicellulase and pectinase gene expression across 1% xylan and 2mM constituent sugars. *Heat map shows differential expression as measured by log_2 fold change between no carbon vs xylan (left column), no carbon vs xylose (middle) and no carbon vs arabinose (right). Log2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar. Genes are ordered by descending value of log_2 fold change between no carbon and xylan.*

Xylose and arabinose play an insignificant role in inducing cellulase gene expression. If we examine the expression of cellulases and LPMOs under these conditions, we observe that a very small number of cellulases are LPMOs are up regulated to a low level.

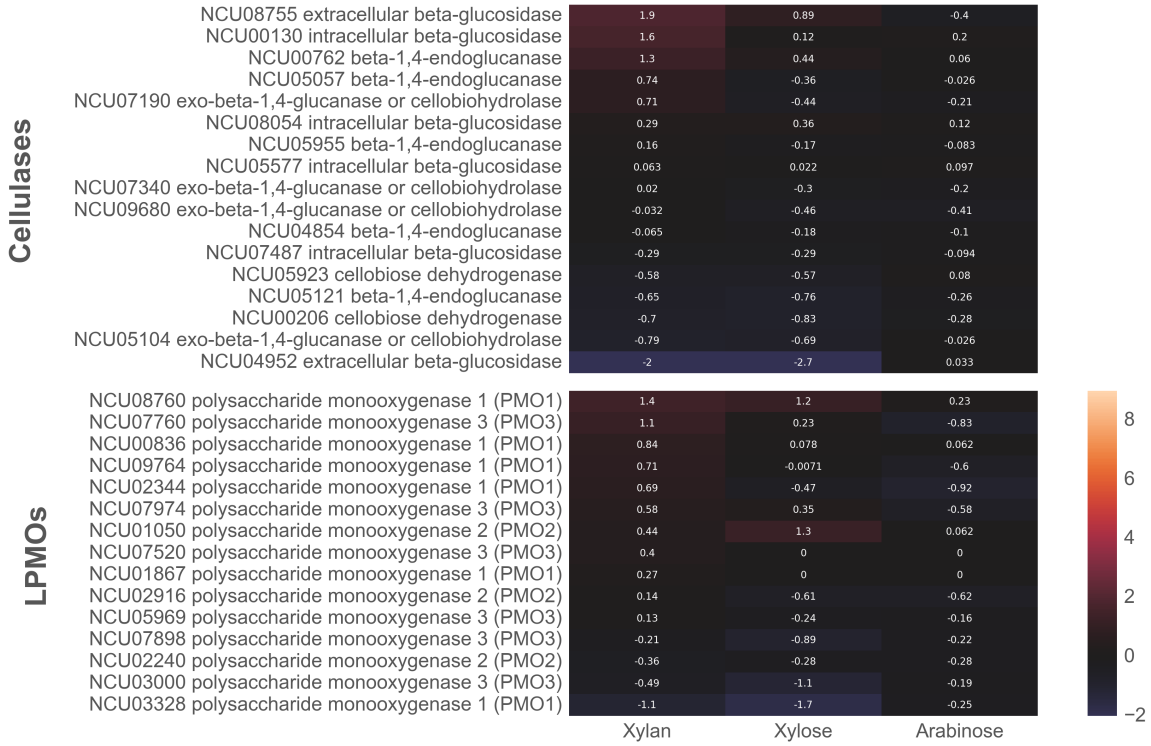


Figure 3-8 Comparison of cellulase gene expression across xylan conditions. *Heat map shows differential expression as measured by log_2 fold change between no carbon vs xylan (left column), no carbon vs xylose (middle) and no carbon vs arabinose (right). Log_2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar. Genes are ordered by descending value of log_2 fold change between no carbon and xylan.*

Pectin and its constituent sugars

Pectin is much more complex polysaccharide as compared to cellulose and xylan. The two main types of pectin are homogalacturonan (HG) and rhamnogalacturonan I (RG1). GalA is the primary sugar that makes up the backbone of HG, while RG1 backbone is composed of alternate GalA and rhamnose residues. The side chains are composed of a variety of sugars including arabinose, xylose, galactose and others. We explored expression of PCWDEs genes after induction by GalA, rhamnose and galactose in comparison to pectin and the two pectic polysaccharides, HG and RG1.

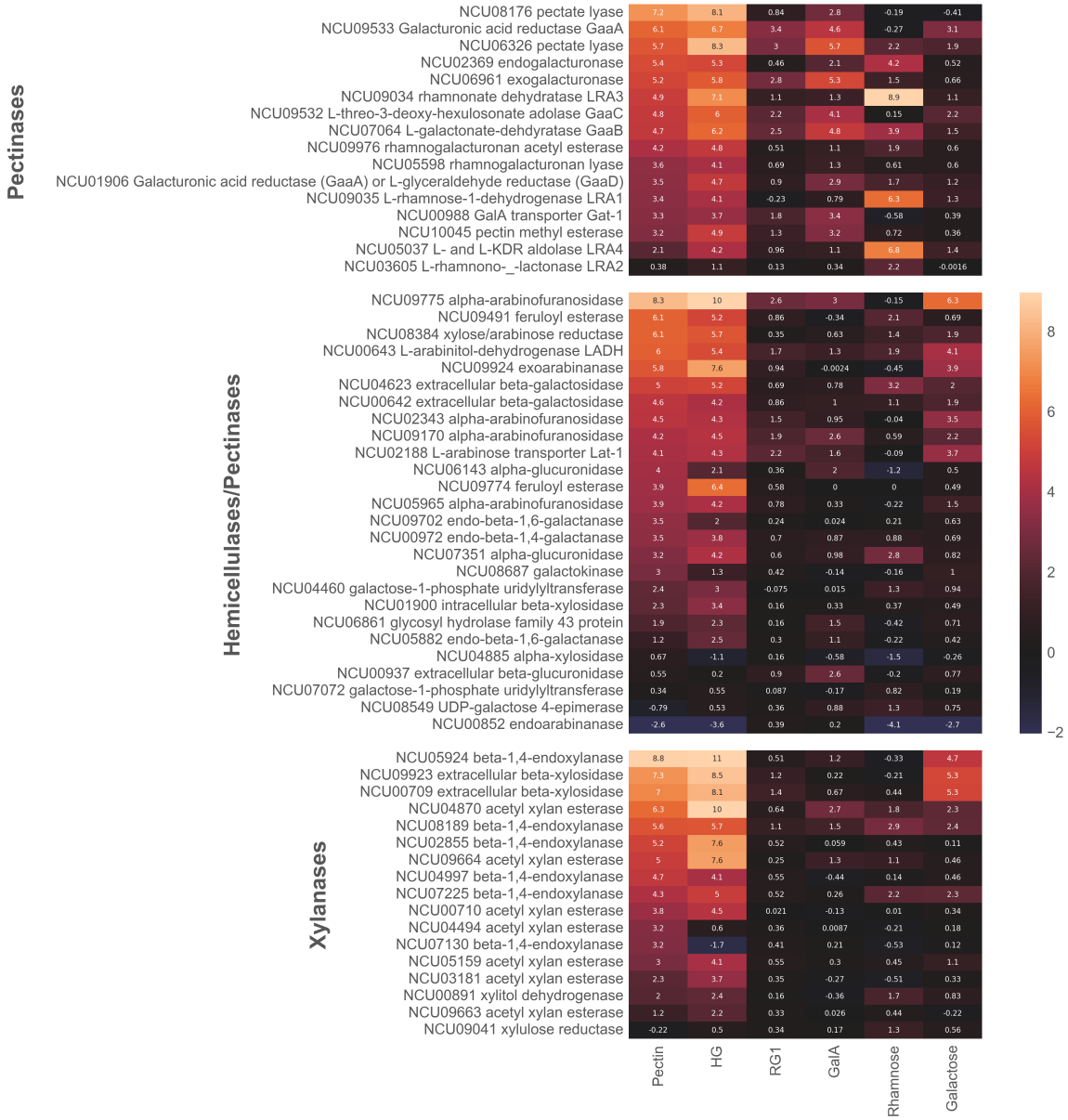


Figure 3-9 Comparison of genes encoding pectinases and hemicellulases across pectinolytic conditions. *Heat maps show differential expression as measured by log_2 fold change between no carbon and 6 pectin related conditions: Pectin, homogalacturonan, rhamnogalacturonan I, galacturonic acid, rhamnose and galactose (left to right). Log_2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar. Genes are ordered by descending value of log_2 fold change between no carbon vs pectin.*

We observed that the majority of genes encoding pectinases are up regulated on pectin were also induced by either GalA or rhamnose. Rhamnose appears to be the major inducer of four rhamnose catabolism enzymes LRA1-4 (NCU09035, NCU03605, NCU09034, NCU05037), as well as the endogalacturonanase (NCU02369), and the GalA catabolism enzyme GaaB (NCU07064). GalA induces the

remaining key pectinase genes including GalA catabolism enzymes GaaA-D (NCU09533, NCU07064, NCU09532, NCU01906). Galactose also plays a minor role in pectin metabolism and induces expression or GaaA-D, LRA1, LRA4, and a pectate lyase (NCU06326) to low levels (Figure 3-9).

A moderate set of hemicellulase and xylanase genes are up regulated by galactose, and not by GalA or rhamnose. These include four endoxylanase genes and the only two β-xylosidases present in the genome. A number of arabinose metabolic genes and β-galactosidases are also up regulated as shown under hemicellulase/pectinase category of Figure 3-9. Interestingly, this very similar set of genes is also up regulated under arabinose induction (Figure 3-7), which indicates that the two sugars might be acting in the same regulatory network.

A large number of cellulase genes were induced under pectin and HG. This was surprising to us as the top expressed cellulase genes under pectin and HG (Figure 3-10) were nearly as highly expressed under cellulolytic conditions (Figure 3-5). These cellulases are also not induced directly by GalA, rhamnose, or galactose, so there must be something else in these substrates that is inducing expression of cellulase genes. We believe that contaminating cellulose is the most likely cause for this induction, as the major sugars of pectin have almost no effect on cellulase gene expression.

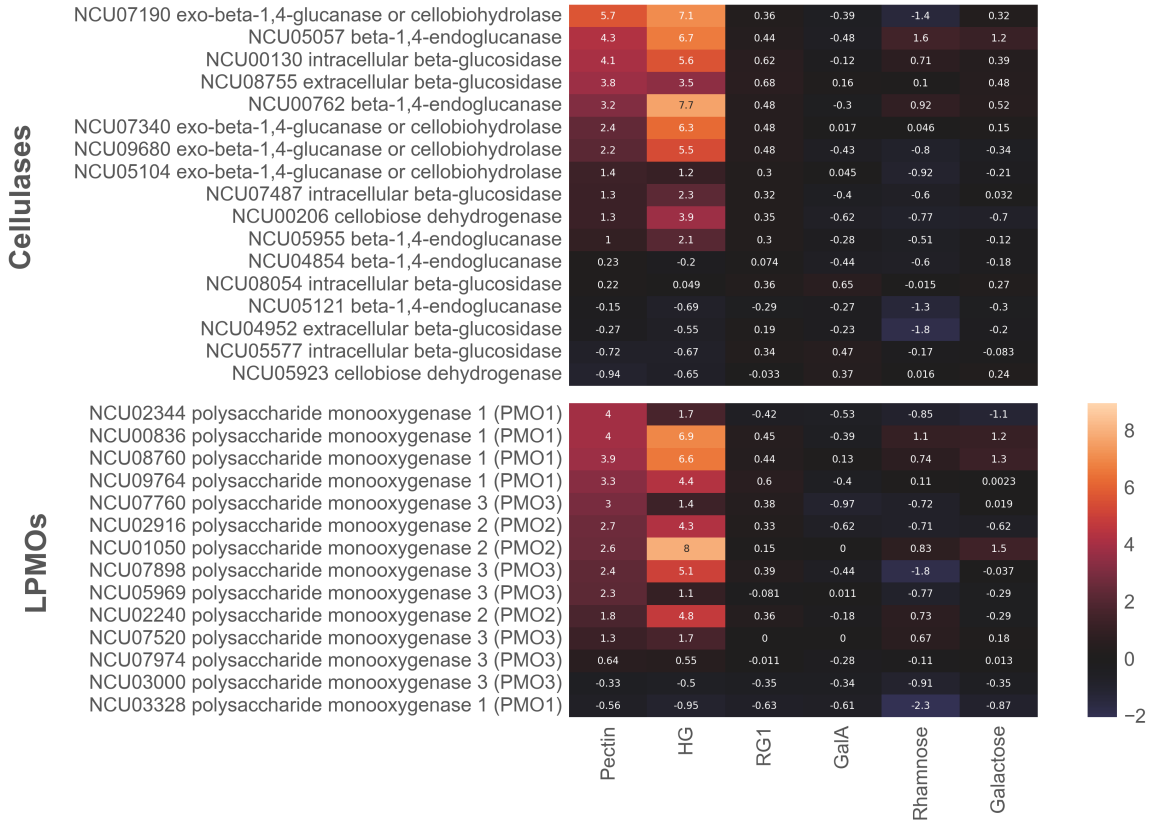


Figure 3-10 Comparison of cellulase gene expression across pectinolytic conditions. *Heat maps show differential expression as measured by log_2 fold change between no carbon and 6 pectin related conditions: Pectin, homogalacturonan, rhamnogalacturonan I, galacturonic acid, rhamnose and galactose (left to right). Log_2 fold change is depicted in the center of each cell, and colored appropriately*

according to the gradient on the color bar. Genes are ordered by descending value of log₂ fold change between no carbon vs pectin.

We also explored expression of PCWDEs genes under our mannan conditions, comparing galactomannan and glucomannan to mannose and mannobiose. Mannans are a common hemicellulose, but when we explore the *N. crassa* genome, there is a paucity of mannan degradation genes. In fact, there are only three: one β-1,4-endomannanase (NCU08412), one intracellular β-mannosidases (NCU00890) and one extracellular β-mannosidase (NCU00985). When we examine expression of these three genes under mannan conditions, we observe that the endomannanase is up regulated across all four conditions, however only 2mM mannose induces expression of the β-mannosidases and only to a low extent (Figure 3-11).

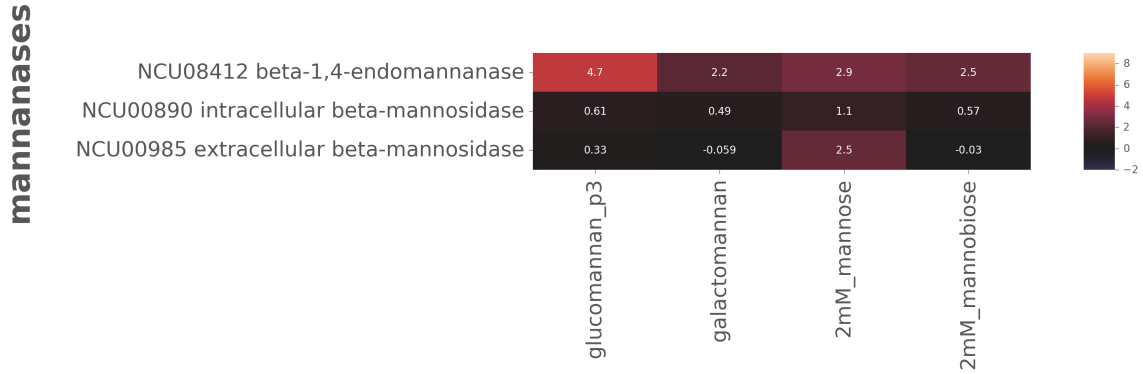


Figure 3-11 Comparison of mannanase gene expression across mannan conditions. *Heat maps show differential expression as measured by log₂ fold change between no carbon and four mannan conditions: 1% galactomannan, 1% glucomannan, 2mM mannose, and 2mM mannobiose (left to right). Log₂ fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar.*

We found that mannose is a weak inducer of other PCWDE genes while mannobiose is mostly a non-inducer. If we examine cellulases, hemicellulases and pectinases gene expression across our mannan conditions we see that certain PCWDEs are up regulated under mannose and our full mannan conditions. Most importantly, mannose itself is promoting this expression, as glucomannan and galactomannan could be compromised with impurities. Mannose induces weakly a very small number of cellulase genes (data not shown), but does induce expression of a large number of hemicellulases and pectinases, albeit not to the extent observed under sugars like xylose, arabinose and galactose (Figure 3-12).

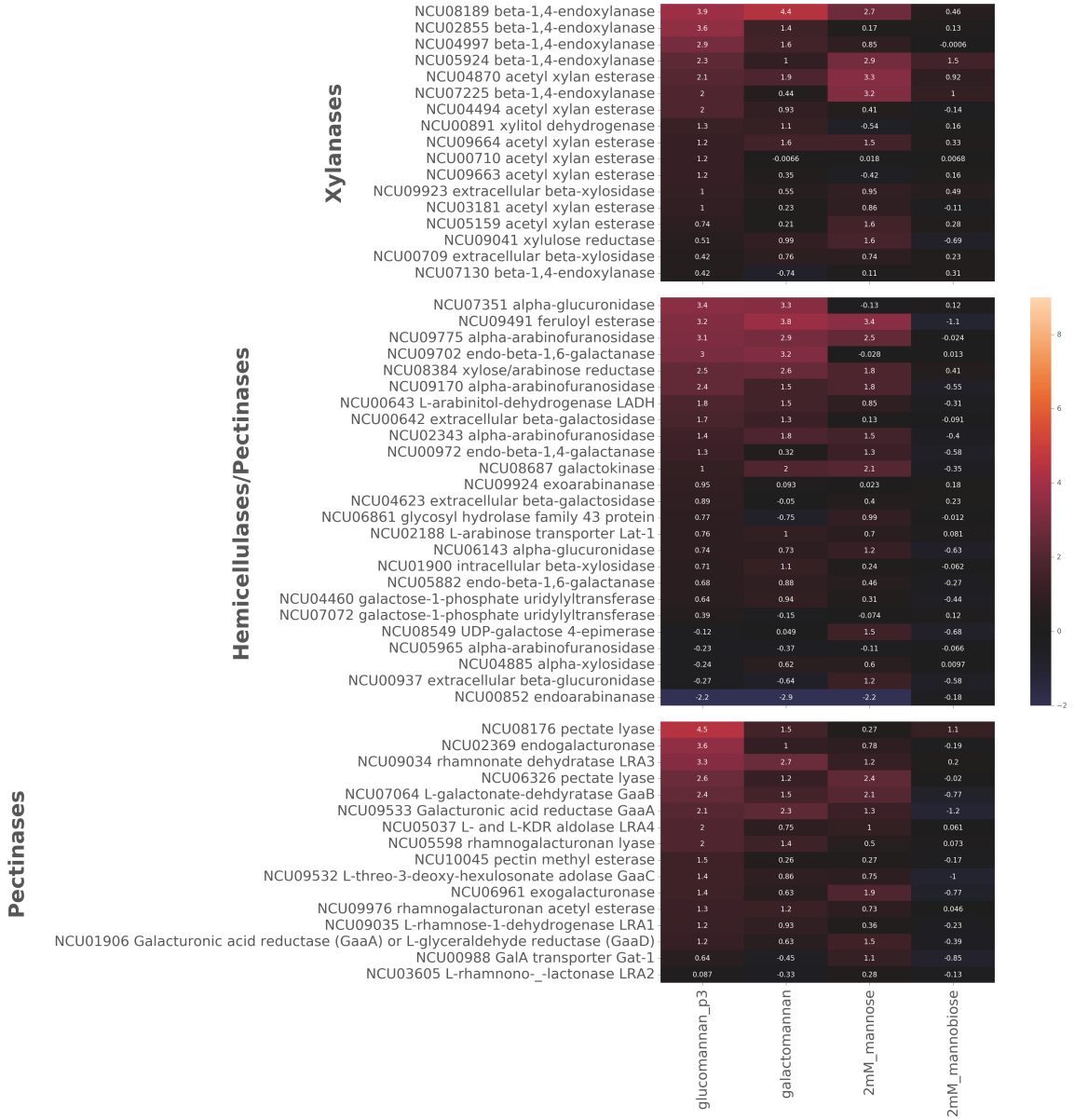


Figure 3-12 Comparison of hemicellulase gene expression across mannan conditions. *Heat maps show differential expression as measured by log_2 fold change between no carbon and four mannan conditions: 1% galactomannan, 1% glucomannan, 2mM mannose, and 2mM mannobiose (left to right). Log_2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar.*

We also explored expression of starch depolymerization genes under starch conditions. Amylose, a helical polymer, and amylopectin, a branched polymer are the two constituents of starch and exist at about 1:4 ratio (125). We were surprised to find that only a portion of the starch depolymerization genes found in the *N. crassa* genome was induced under amylose and amylopectin. *N. crassa* appears to have a number of α-glucosidase genes that are not induced by any of our four starch conditions. In fact, expression of these genes under these conditions generally falls

below expression under no carbon conditions. These genes are therefore likely unimportant for degradation of these starch molecules.

We found that maltose and trehalose induce a very similar subset of genes as amylopectin, indicating that both disaccharides signal for induction of the same starch degrading genes. Maltose seems to have slightly higher induction than trehalose, but the subset of genes induces are practically the same, and also identical to the genes observed under amylopectin (Figure 3-13). These data indicate that these starch components likely act upon the same regulatory network.

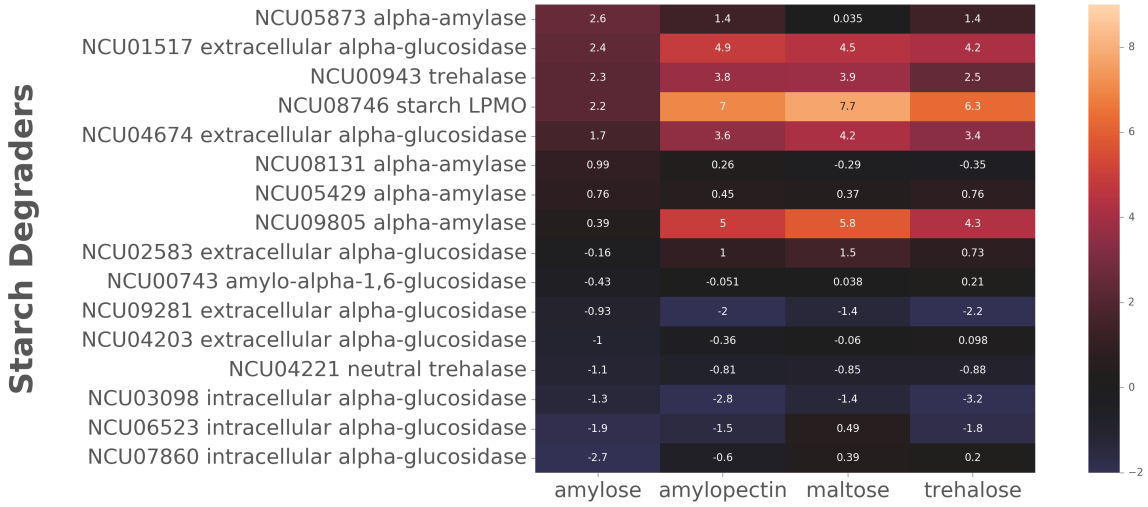


Figure 3-13 Comparison of expression levels of starch degrading genes under starch related conditions. *Heat maps show differential expression as measured by log_2 fold change between no carbon and four starch related conditions: 1% amylose, 1% amylopectin, 2mM maltose, and 2mM trehalose (left to right). Log_2 fold change is depicted in the center of each cell, and colored appropriately according to the gradient on the color bar. Genes are ordered by descending value of log_2 fold change between no carbon vs amylose.*

Summary

Comparisons of PCWDEs gene expression levels between complex polysaccharides and their constituent sugars, informs us on a several key concepts pertaining to PCWDE production in *N. crassa.*

We show that monosaccharides and disaccharides individually can induce unique sets of PCWDEs, but multiple sugars from the same polysaccharide are generally required to fully induce the complete set of genes required to depolymerize that complex polysaccharide. Cellobiose induces a set of cellulase genes comparable to the set induced on Avicel. Xylose and arabinose together induce a set of xylanase genes comparable to xylanases induced by xylan. GalA and rhamnose induce a set of pectinases comparable to pectinases induced on pectin and HG.

Some sugars induce the same set of PCWDEs, which indicate that they likely signal the same transcription factor and its regulatory network. We observe that

galactose and arabinose both induce the same set of hemicellulase and pectinase and xylanase genes. Trehalose and maltose induce expression of the same five starch degradation genes.

When there is an induction of cellulase genes, there is also induction of a number of pectinase and hemicellulose genes, but the opposite is not true. Even under 2mM cellobiose, we see moderately high expression of a number of xylanases, hemicellulases, and pectinases. We can conclude that this induction is not likely due to impurities in cellobiose. However, we only observe very minor expression of cellulase genes under sugars sugar conditions derived from pectin/hemicellulose. These data indicate that cellulose and its derivatives are a more important signal to *N. crassa* than other polysaccharides.

Sugars signal for more than the depolymerization of its parent polymer. Cellobiose exemplifies this concept as it induces expression of pectinases and hemicellulases. We also see that xylose induces a number of arabinases, galatosidases and pectinases in addition to xylanases. Although not as strong inducers as xylose and cellobiose, arabinose and galactose display a similar narrative; both those sugars induce PCWDEs responsible for cleaving sugars absent from arabinan and galactan . Finally, mannose weakly induces a small set of PCWDEs genes including cellulases, hemicellulases and pectinases.

3.4. Novel candidate transcription factors involved in carbon metabolism and PCWDE production

3.4.1. *Introduction*

One of the major objectives of the transcriptome profiling was to understand the transcriptional regulation behind PCWDE production in *N. crassa.* The key components of this process are the TFs that positively regulate the various subsets of PCWDEs. In the recent years, a number of TFs have been discovered that regulate the catabolism of specific polymers in the plant cell wall. Each major category of polysaccharides seems to have its own transcription factors that are active when that polysaccharide is the main carbon source present. These TFs induce expression of PCWDEs and intracellular catabolic genes responsible for degradation and utilization of the polysaccharide. TFs that fit this description have been identified in a number of different ascomycete fungi and many are well conserved across genera. A summary of some of these TFs is presented in the introductory chapter (see section 1.6).

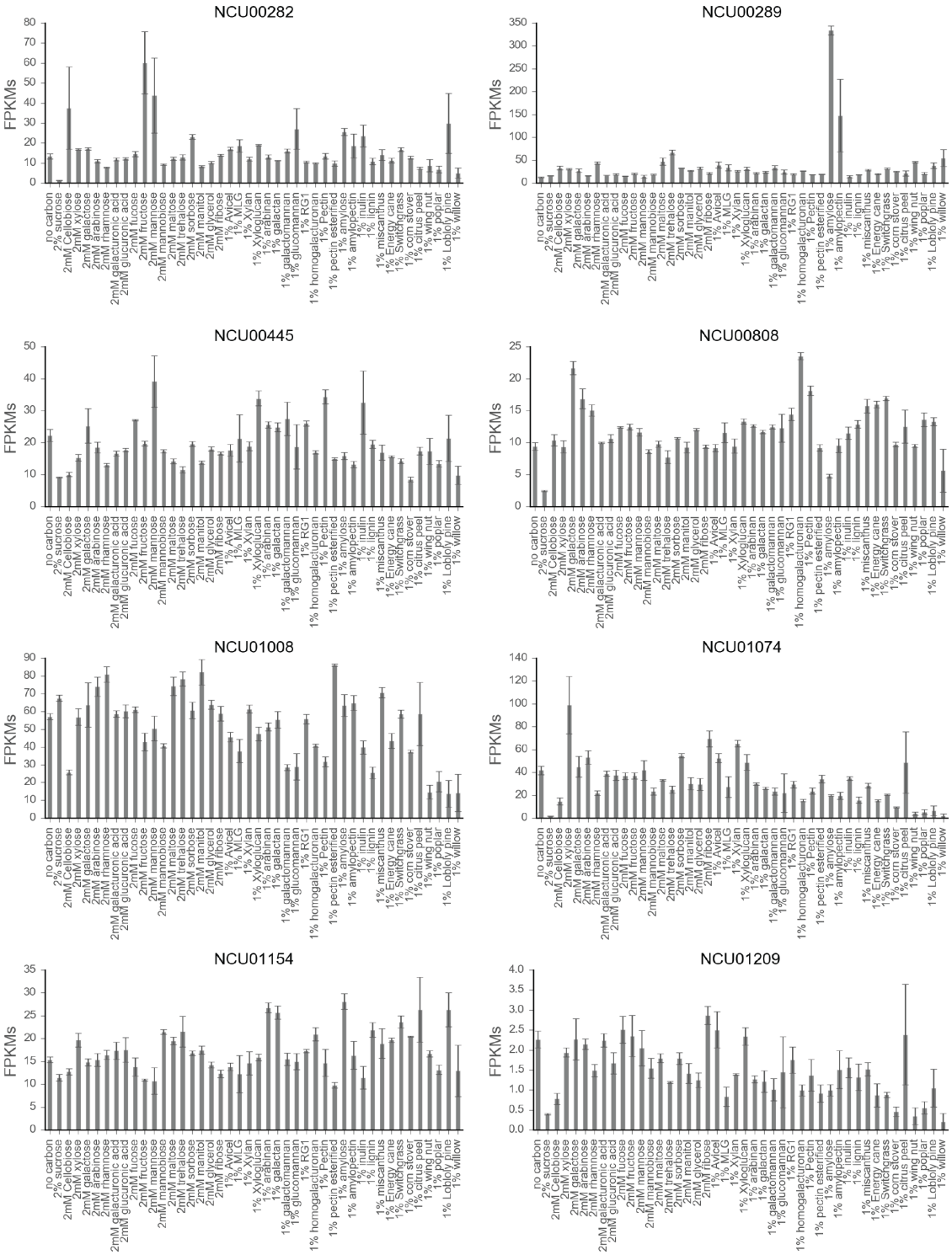
At the time that this project was initiated, the major carbon metabolic TFs discovered in *N. crassa* were CLR-1, CLR-2, XLR-1. These TFs positively regulate the majority of cellulases and xylanases genes along with a number of non-xylan hemicellulose genes (27, 126). We hypothesized that there were additional transcription factors responsible for regulating pectin metabolism. *N. crassa* grows well on pectin and has a suite of pectinases with no described regulators. We were also unsure whether regulation of PCWDEs had additional TFs that worked upstream, downstream or independently of CLRs and XLR-1. Thus we searched for

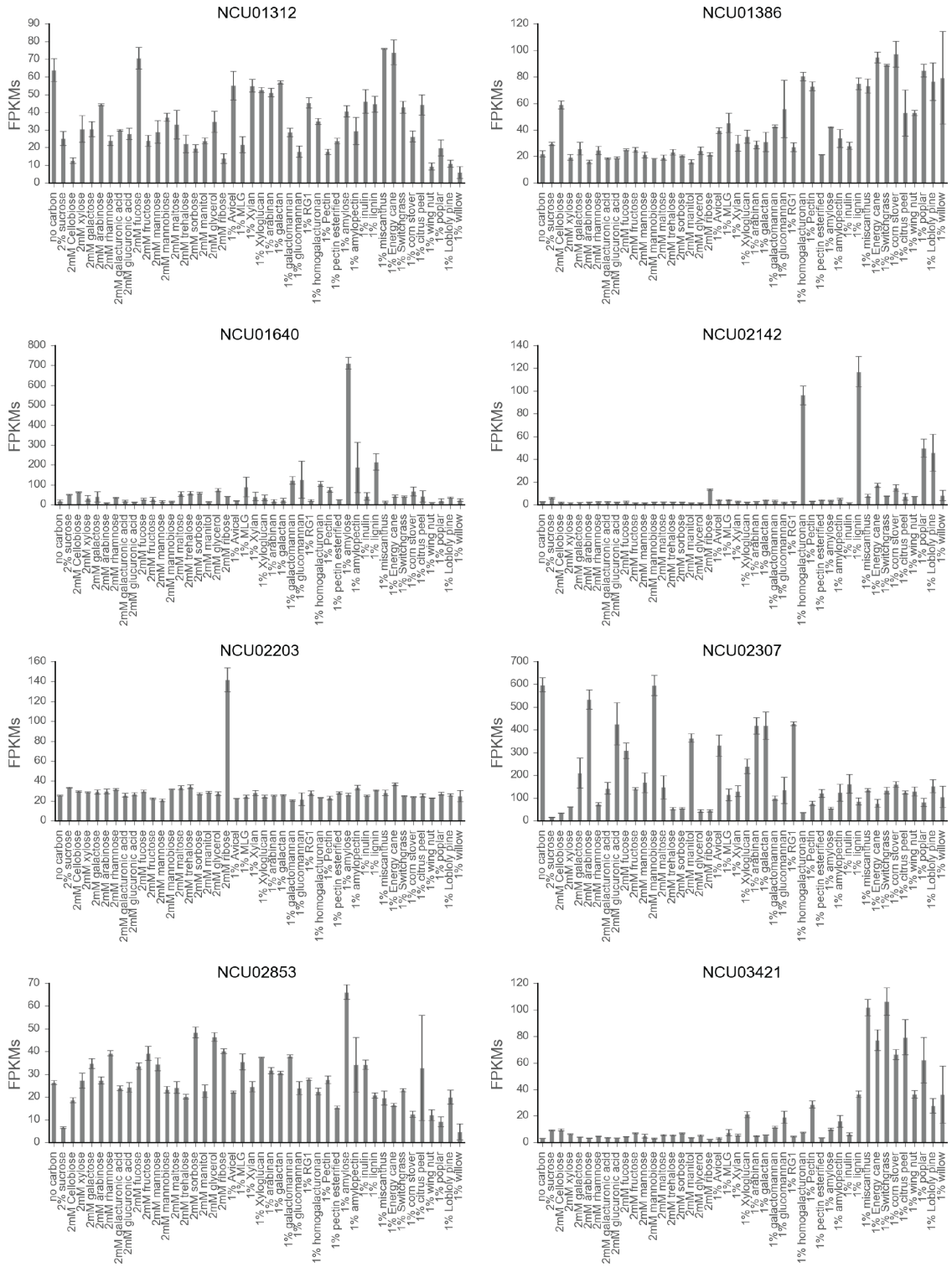
TFs that might be regulating additional aspects of cellulose and hemicellulose metabolism as well as pectin metabolism.

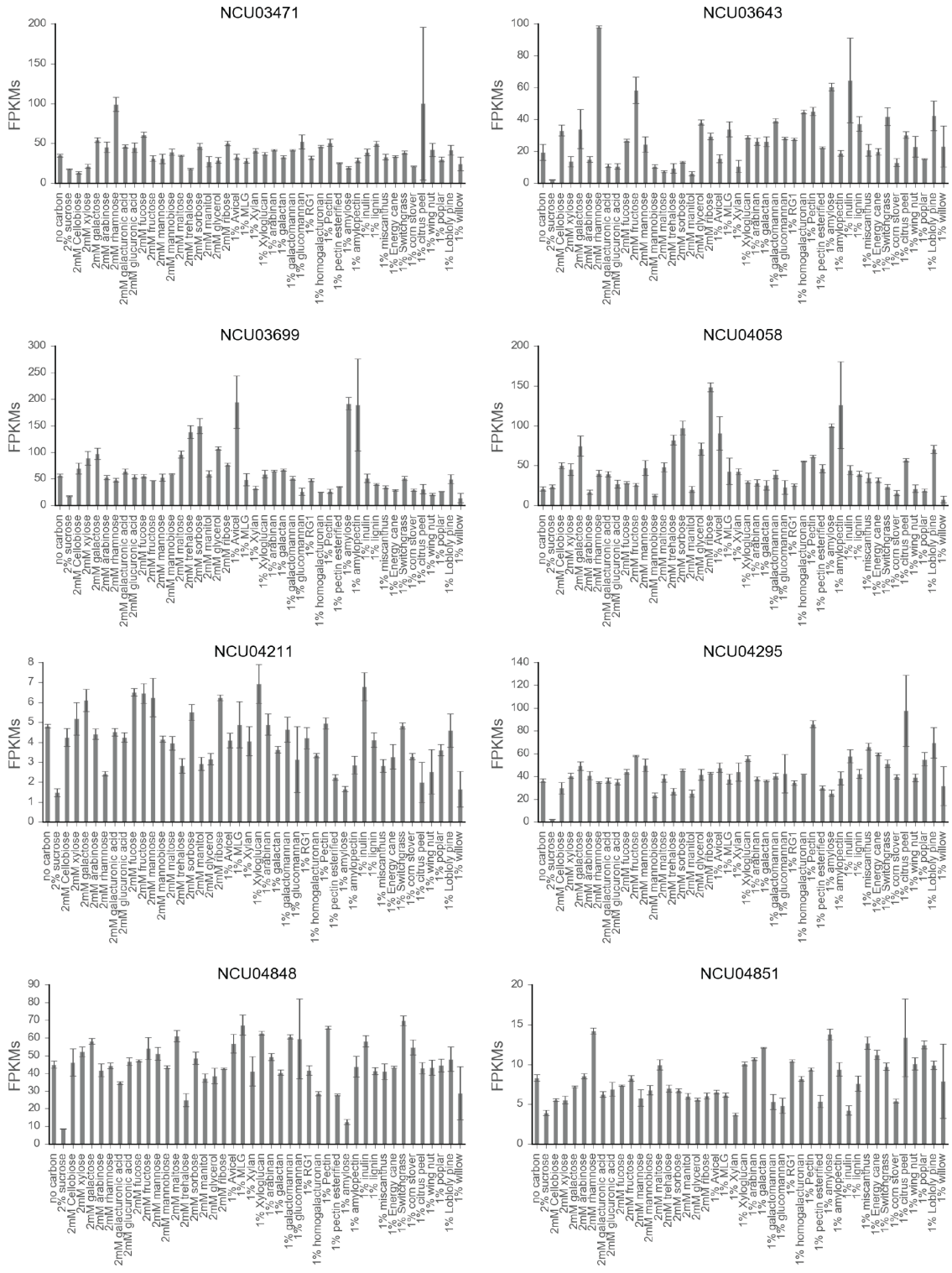
Our strategy consisted of systematic analysis of transcriptomic profiles of known proteins with DNA binding domain. We compiled this list of proteins from various sources including Weiruch et al 2014 (127) and collaborators of that project. We used hierarchical clustering to find TFs with similar expression profiles to PCWDEs. *Clr-1, clr-2 and xlr-1* also had shared expression profile characteristics that we used as criteria to search for novel candidates. For instance, those canonical TFs are down regulated on sucrose, presumably due to CCR, and up regulated on the specific polysaccharides that they degrade. We searched for TFs that had reduced expression on 2% sucrose and increased expression on alternative carbon sources.

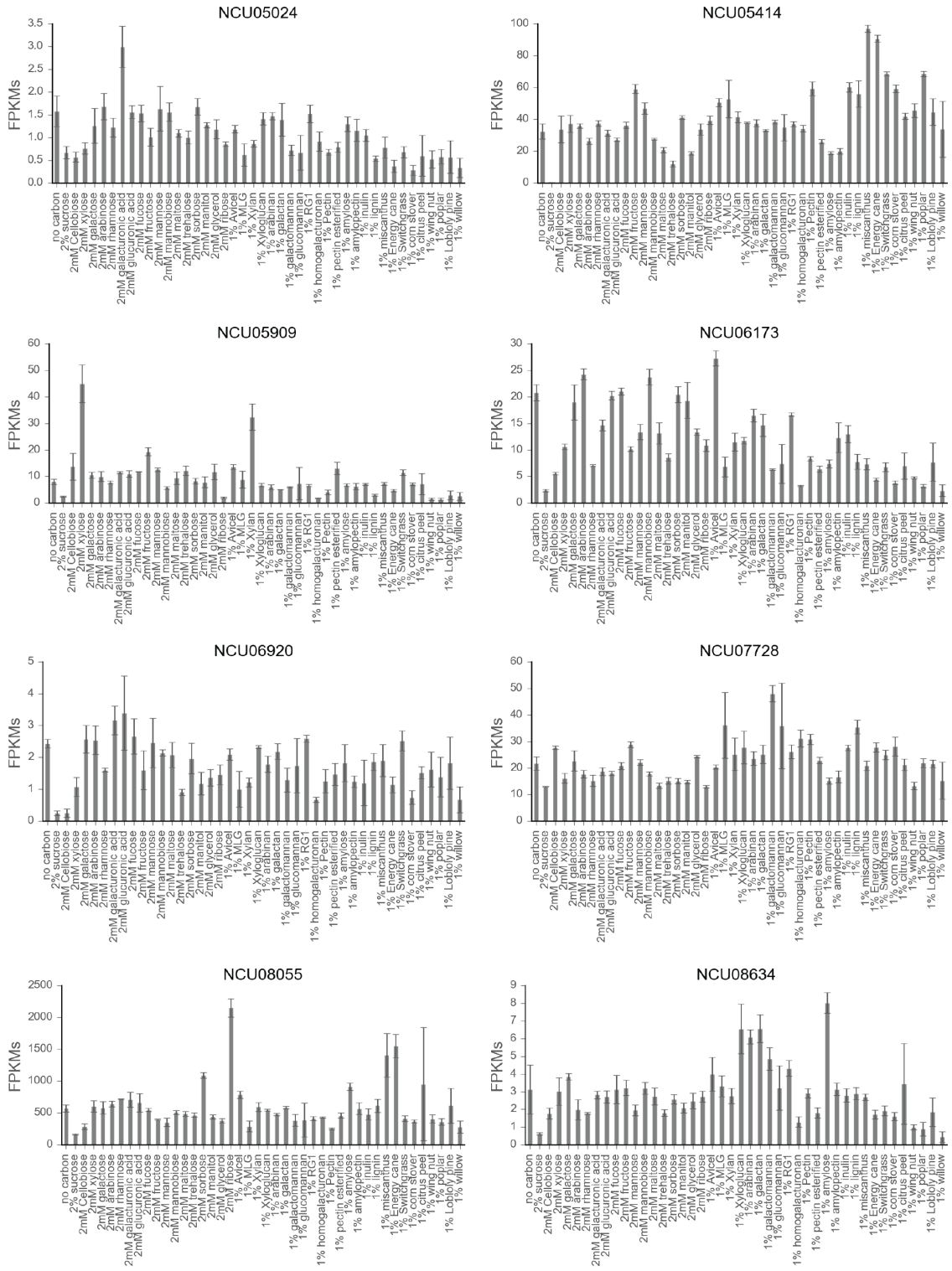
3.4.2. ***Summary of candidate transcription factors found***

We found 37 TF candidates that were up regulated across a number of different nutrient conditions. Displayed below are their expression profiles across our carbon conditions (Figure 3-14). The TFs that are annotated are NCU00289 (*tah-1*, "tall aerial hyphae 1"), NCU01008 (U3 small nucleolar ribonucleoprotein IMP3), NCU01154 (*sub-1,* "submerged protoperithecia-1"), NCU01312 (*rca-1,* "regulator of conidiation in Aspergillus 1), NCU03643 ("cutinase TF 1 β"), NCU07728 ("siderophore regulation, sre"), NCU08055 ("b-zip transcription factor IDI4"). The remaining candidates are not annotated. Further exploration of the function of these TFs is continued in Chapter 4.


NCU00282
NCU00289
NCU00445
NCU00808
NCU01008
NCU01074
NCU01154
NCU01209
FPKMs
no carbon
2% sucrose
2mM Cellobiose
2mM xylose
2mM galactose
2mM arabinose
2mM mannose
2mM galacturonic acid
2mM glucuronic acid
2mM fucose
2mM fructose
2mM mannose
2mM mannobiose
2mM maltose
2mM trehalose
2mM sorbose
2mM mannitol
2mM glycerol
2mM ribose
1% Avicel
1% MLG
1% Xylan
1% Xyloglucan
1% arabinan
1% galactan
1% galactomannan
1% glucomannan
1% RG1
1% homogalacturonan
1% Pectin
1% pectin esterified
1% amylose
1% amylopectin
1% inulin
1% lignin
1% miscanthus
1% Energy cane
1% Switchgrass
1% corn stover
1% citrus peel
1% wing nut
1% poplar
1% Lobolly pine
1% willow


NCU01312
NCU01386
NCU01640
NCU02142
NCU02203
NCU02307
NCU02853
NCU03421
FPKMs
no carbon
2% sucrose
2mM Cellobiose
2mM xylose
2mM galactose
2mM arabinose
2mM rhamnose
2mM galacturonic acid
2mM glucuronic acid
2mM fucose
2mM fructose
2mM mannose
2mM mannobiose
2mM maltose
2mM trehalose
2mM sorbose
2mM mannitol
2mM glycerol
2mM ribose
1% Avicel
1% MLG
1% Xylan
1% Xyloglucan
1% arabinan
1% galactan
1% galactomannan
1% glucomannan
1% RG1
1% homogalacturonan
1% Pectin
1% pectin esterified
1% amylose
1% amylopectin
1% inulin
1% lignin
1% miscanthus
1% Energy cane
1% Switchgrass
1% corn stover
1% citrus peel
1% wing nut
1% poplar
1% Loblolly pine
1% willow


NCU03471
NCU03643
NCU03699
NCU04058
NCU04211
NCU04295
NCU04848
NCU04851
FPKMs
no carbon
2% sucrose
2mM Cellobiose
2mM xylose
2mM galactose
2mM arabinose
2mM rhamnose
2mM galacturonic acid
2mM glucuronic acid
2mM fucose
2mM fructose
2mM mannose
2mM mannobiose
2mM maltose
2mM trehalose
2mM sorbose
2mM mannitol
2mM glycerol
2mM ribose
1% Avicel
1% MLG
1% Xylan
1% Xyloglucan
1% arabinan
1% galactan
1% galactomannan
1% glucomannan
1% RG1
1% homogalacturonan
1% Pectin
1% pectin esterified
1% amylose
1% amylopectin
1% inulin
1% lignin
1% miscanthus
1% Energy cane
1% Switchgrass
1% corn stover
1% citrus peel
1% wing nut
1% poplar
1% Lobolly pine
1% willow


NCU05024
NCU05414
NCU05909
NCU06173
NCU06920
NCU07728
NCU08055
NCU08634
FPKMs
no carbon
2% sucrose
2mM Cellobiose
2mM xylose
2mM galactose
2mM arabinose
2mM rhamnose
2mM galacturonic acid
2mM glucuronic acid
2mM fucose
2mM fructose
2mM mannose
2mM mannobiose
2mM maltose
2mM trehalose
2mM sorbose
2mM mannitol
2mM glycerol
2mM ribose
1% Avicel
1% MLG
1% Xylan
1% Xyloglucan
1% arabinan
1% galactan
1% galactomannan
1% glucomannan
1% RG1
1% homogalacturonan
1% Pectin
1% pectin esterified
1% amylopectin
1% inulin
1% lignin
1% miscanthus
1% Energy cane
1% Switchgrass
1% corn stover
1% citrus peel
1% wing nut
1% poplar
1% Lobolly Pine
1% willow

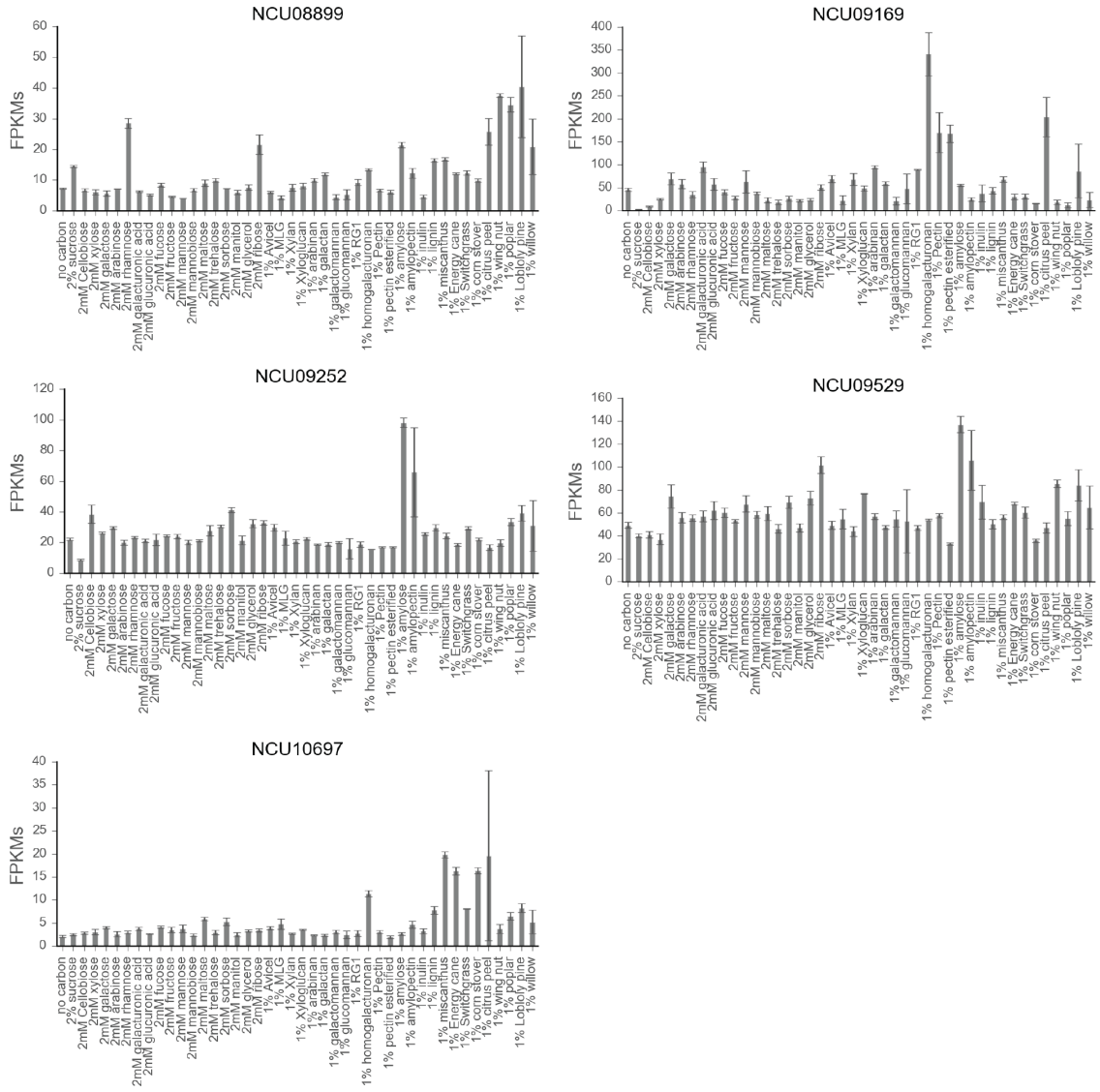


Figure 3-14 Expression profiles of candidate transcription factors found by RNA profiling. *Expression measured by FPKMs averaged from biological triplicates. Error bar represents standard deviation from biological triplicates.*

3.5. Analysis of *N. crassa* MFS sugar transporters provides a number of candidates for further study

3.5.1. *Introduction*

Transporters are an integral part of plant cell wall deconstruction. These proteins bring sugars across the cell membrane and into the cell to be catabolized for energy. Additionally, as we have shown with *N. crassa*, once these sugars are inside the cell they are signal transducers for plant cell wall degradation. Sugars activate transcription factors, which can then promote or repress expression of PCWDEs depending on their concentration. The canonical model for this process is

demonstrated by cellobiose import, where low levels of cellobiose can induce expression of cellulases by activation of CLR-1 and CLR-2 (27, 43).

Filamentous fungi have an enormous diversity of membrane transporters, but sugar transporters so far have been limited to the Major facilitator superfamily (MFS). When we examined the *N. crassa* genome for transporters using the transporter classification database (TCDB) (http://www.tcb.org/), we found 703 transporters in total. 133 of these transporters are a part of the MFS family, and all of the sugars transporters annotated to date fall into this group. When we built a maximum likely hood tree using protein sequences of all the MFS transporters, we noticed that all of the known sugar transporters fell within a subclade within the tree (Figure 3-15). This clade is labeled in purple, and within is an alternate purple color tone representing the clade in which the cellodextrin transporters, CDT-1, CDT-2, and CDT-3 or CBT-1 (128) belong to.

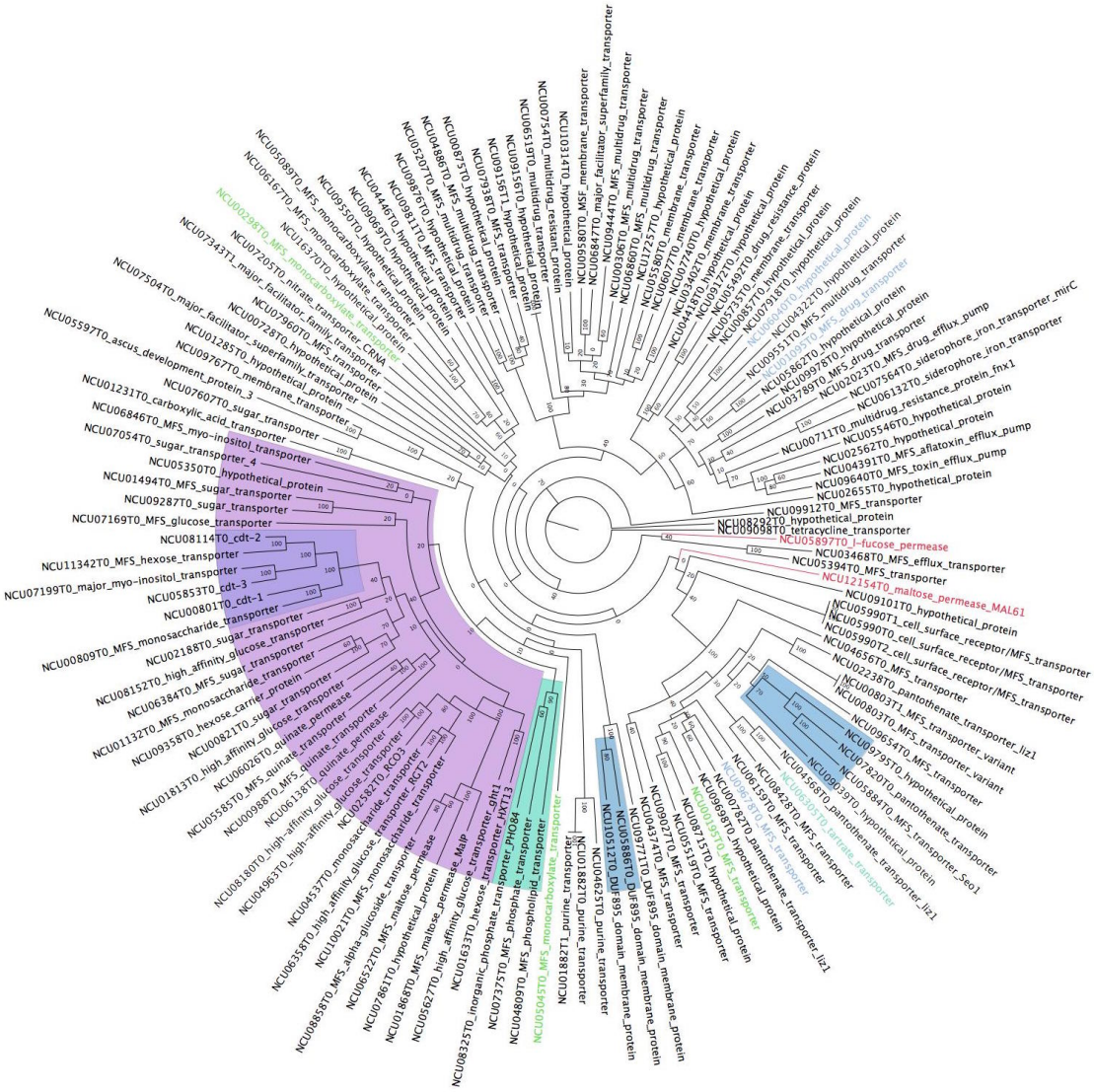


Figure 3-15 Maximum likelihoood tree of all MFS transporter genes predicted in the *N. crassa* genome. *Maximum likelihood tree built with RaxML*(69) *with 10 bootstrap replicates (additional replicates were not conducted due to extended*

computational time). Bootstrap values denoted in the nodes of each branch. Tips are labeled with Gene ID and Broad v12 annotation. Highlighted clades contain transporters of interest. Purple highlights sugar transporters. Green represents possible involvement in phosphate acquisition, and blue for sulfur acquisition. Red highlighted genes represent putative sugar transporters outside of the purple clade.

The majority of the transporters in the MFS tree are not well annotated. Many of them do have annotations based on homology to other transporters, but often times these are misleading and incorrect. One example is a transporter previously annotated as a quinate transporter due to its homology to a biochemically validated quinate transporter, but recent research has shown that it transports GalA (129). This aspect has been a common theme for many transporters that have been re-annotated due to findings from biological experiments (130–133). These studies have been critical in developing organisms into cell factories. *S. cerevisiae* has especially been a popular target for engineering in attempts to allow strains to utilize alternative sugars like xylose, GalA and other sugars incompatible with native catabolic pathways (134, 135). One major bottleneck is getting enough of a target sugar imported into the cell. Due to the poor annotation of alternative sugar transporters in filamentous fungi, it can be difficult to find enough transporters for heterologous expression in host systems.

We have derived a list of candidate transporters that are likely critical for transport of a number of alternative sugars. These sugars are many of the un-annotated transporters part of the sugar transporter clade within MFS group. Examining the transcription profile of these candidates transporter has provided us with further information regarding the specific sugars they might be active on. We find a number of candidate general sugar transporters as well as a few transporters only up regulated on very specific 2mM sugar conditions. Additionally we find transporters specifically up regulated under sulfur and phosphate starvation conditions, indicating that they may be involved in acquisition of those nutrients.

3.5.2. ***Results***

When we examined the subclade of MFS sugar transporters that contain all the known sugar transporters we observed that the annotation for many of these transporters had new information published regarding their function. We re-annotated some of the transporters found in the recent literature and demarcated them on the MFS transporter tree (Figure 3-16). We also attached a small description to each candidate transporter that display up regulation on specific conditions. Many are further explored in the upcoming section. Of the transporters that do not have a full annotation, we have divided them into general sugar transporters and specific sugar transporters based on their transcriptional profile. Those up regulated on few sugars have been categorized as specific sugar transporters and those expressed at high levels across many sugar conditions we have deemed general transporters. The next sections explore the transcriptional profile of the more highly expressed un-annotated putative sugar transporters.

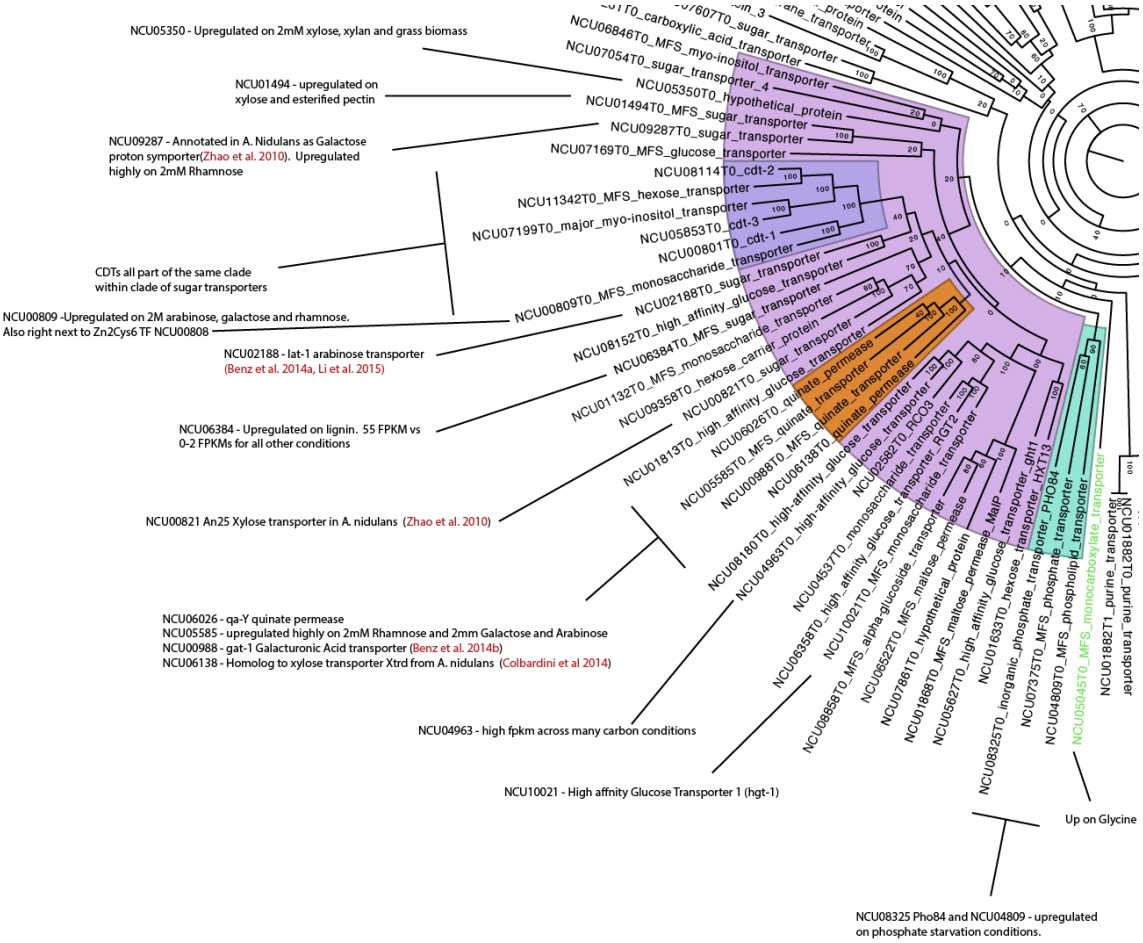


Figure 3-16 MFS transporter sugar sub-clade. *This is a portion of the full N. crassa MFS transporter tree that contains all of the annotated mono and disaccharide transporters, denoted in purple highlight. Within this group, a secondary purple highlight demarcates the clade in which the all known cellodextrin transporters belong to. The orange highlighted clade contains transporters previously annotated as quinate permeases due to their homology with NCU06026, but are likely mis-annotated. Green highlight denotes clades containing pho-84, a phosphate transporter, which also contains another transporter highly up regulated under phosphate starvation conditions.*

We found eight un-annotated MFS transporters that have transcription profiles indicating specificity for unique monosaccharides. Their expression profiles are shown below (Figure 3-17). Most are up regulated on hemicellulose/pectin conditions indicating that they may be transporting sugars cleaved from those polysaccharides. Transporter is up regulated on lignin and full plant biomasses, an indication that it may be involved in transporting a compound present in plant biomass but not in the complex polysaccharides that we tested. Several more transporters to note are NCU01132, NCU05585 and NCU04537. These three transporters have much higher expression levels compared to the others and thus

may play more important roles in sugar uptake.

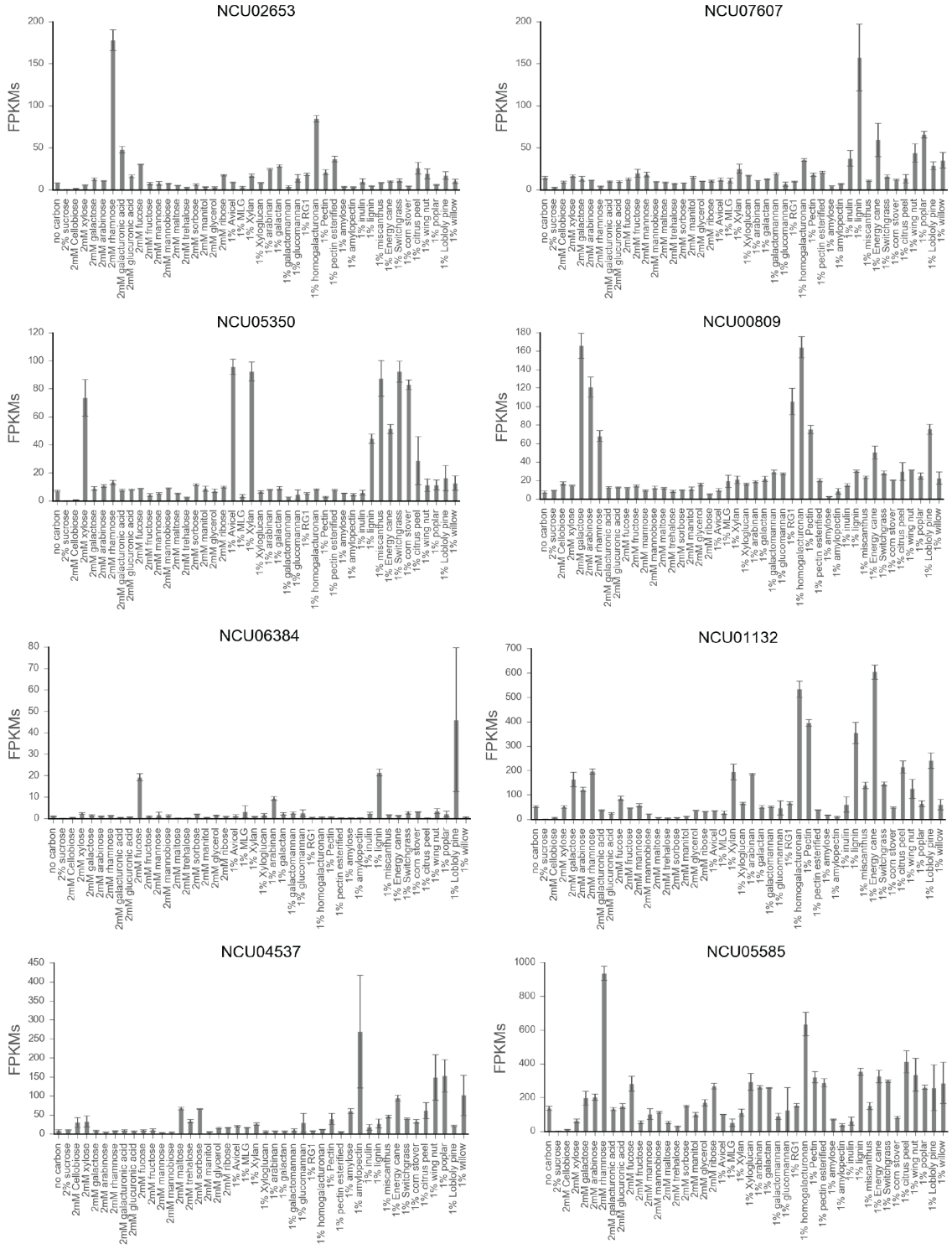


Figure 3-17 Transcriptional profiles of MFS sugar transporters up regulated across conditions. *Expression measured by FPKMs averaged from biological triplicates. Error bar represents standard deviation from biological triplicates.*

We found eight transporters of interest whose transcription profiles indicate that they may be general sugar transporters. Two transporters, NCU01231 and NCU01633, have a magnitude higher level of expression level than our other candidate transporters. We believe that the level and pattern of expression of these two genes indicate that they are likely more important to general sugar acquisition than the others (Figure 3-18). Several additional un-annotated sugar transporters are not shown here due to their low expression levels.

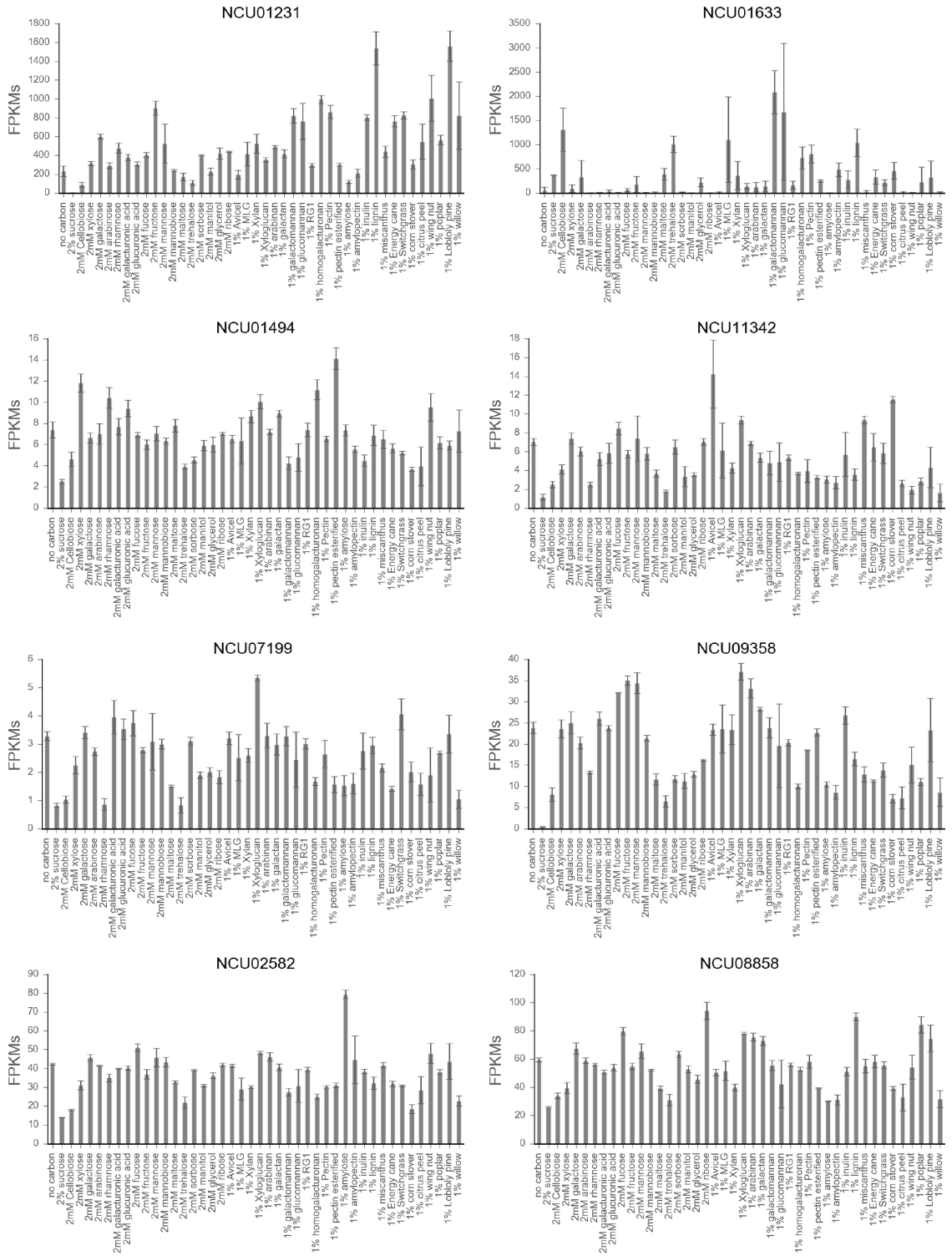


Figure 3-18 Transcription profiles of candidate general sugar transporters.
Expression measured by FPKMs averaged from biological triplicates. Error bar represents standard deviation from biological triplicates.

We found additional un-annotated transporters in the MFS group that are up regulated on our phosphate and sulfur conditions. These transporters are not within the sugar transporter subclade. We believe that they may be involved in the acquisition of phosphate and sulfur due to their transcriptional profiles. Transporters solely up on sulfur starvation include NCU06040, NCU07820, NCU09039, NCU10512. These transporters are further discussed in section 5.8 in regard to sulfur regulator CYS-3. We additionally found one transporter (NCU09767) that was up regulated on no phosphate, no sulfur and no nitrogen.

Discussion

From our cross condition RNAseq, we found a number of un-annotated MFS transporters with interesting with transcriptional profiles that indicate targeting of specific sugars. We first found that all known sugar transporters belong to monophyletic group within the tree of MFS transporters. Thus it is likely that other transporters within this MFS group are also sugar transporters. All of our candidate transporters belong to this clade, and many have transcriptional profiles indicating that they are transporting specific sugars. Identifying these candidate sugar transporters is an important first step in annotating transporters that will help us better understand how plant cell wall catabolism is regulated. Additionally, we have additional data that indicate that transporters play a much larger role in carbon catabolite repression than previously thought (see chapter 5). Understanding which transporters act on which sugars would help dissect the mechanism of this process. Additionally, sugar transporters are very important for engineering organisms to be able to utilize sugars from pectin and hemicellulose.

The MFS sugar transport clade that we found has also aided our analysis of data presented throughout this dissertation. We explore the expression of these transporters in response to various carbon, nitrogen, phosphate, and sulfur sources and use the information to generate new hypotheses regarding function of these proteins.

3.6. Nitrogen, phosphate and sulfur profiling

Nitrogen, phosphate and sulfur acquisition are also critical for cell survival. We profiled the transcriptome of *N. crassa* exposed to a diverse set of nitrogen sources including preferred nitrogen sources like glutamine, glutamate and ammonium to non-preferred sources such as nitrate or BSA. We also included a number of conditions where unique amino acids were the sole nitrogen source. The sulfur and phosphate conditions we tested were more limited. We included no, low and high concentrations for sulfur and phosphate conditions (see methods section 3.2.1).

We wanted to determine how carbon metabolism is affected by the metabolism of these other substrates. The basic strategy we implemented examined whether PCWDEs and carbon metabolic genes expression changed in these various conditions. Sucrose at 2% w/v was the carbon source for all N, P, S conditions. Thus carbon metabolic genes that showed a change in expression must also bypass

CCR. Genes that fit this description are likely regulated by the same machinery that govern N, P, S metabolism.

The N, P, S data set will also be a useful reference for future researchers studying metabolism in fungi. The information regarding N, P, S metabolism is long out dated, and many important details remain unresolved. Most of the work regarding metabolism of these nutrients used genetics experiments to decipher the pathways of key metabolic genes and the transcription factors that regulate them. No functional genomics has been conducted to fully understand the transcriptional regulators discovered in these studies.

3.6.1. *N. crassa response to phosphate starvation*

Introduction

N. crassa uses a simple system to regulate phosphate uptake and acquisition. A key transcription factor responsible for regulating phosphate uptake in *N. crassa* was discovered late 1960's and early 1970's by a research group at the University of Tokyo, Japan (136–139). These researchers originally screened a set of UV mutants for strains unable to utilize nucleotides as their sole source of phosphorous and discovered two genetic loci of importance. The first locus *nuc-1* was shown to encode a transcription factor responsible for regulating genes involved in phosphorous acquisition (52). These include several phosphatases (*pho-2* and *3*) and an inorganic phosphate transporter (*pho -4*). The other loci found encoded three more proteins upstream of *nuc-1*. Two regulatory proteins, PREG and PGOV form a complex to prevent NUC-1 translocation into the nucleus in the presence of phosphate (140). In absence of phosphate NUC-2, an ankyrin repeat protein upstream of *preg/pgov* inhibits PREG-PGOV complex function thus allowing NUC-1 to enter the nucleus (52, 138, 140, 141). After publication of these early studies further details regarding full function of NUC-1 and other TFs with regard to phosphate starvation has yet to be explored. We used RNAseq to try to gain a better understanding of the NUC-1 regulon and other TFs that might be involved in phosphate acquisition.

Results

N. crassa responds dramatically to phosphate limitation

Gene expression changes dramatically between presence and absence of phosphate, while expression changes very little between high phosphate and low phosphate conditions. Between high and low phosphate, few genes showed significant differential expression (Data not shown). ~1.3 fold change was the greatest fold change for any gene. These data suggest that our low phosphate conditions (50mg/L) contained more than enough phosphate for normal growth after 4 hours, and our high phosphate conditions (1g/L) contained phosphate far beyond what is necessary to sustain growth for 4 hours. In contrast, *N. crassa* implements a dramatic transcriptional response when completely starved of phosphate. Differential expression between no phosphate condition as compared to high phosphate conditions shows 61% of genes are differentially expressed. Lack of phosphate causes cells to dramatically alter its expression profile. When we

examine top differentially expressed genes we see many phosphate acquisition genes are present. The top two genes are an alkaline phosphatase (NCU01376) pho-2 and acid phosphatase (NCU08643) *pho-3*. Within the top 10 differentially expressed genes are the inorganic phosphate transporters *pho-4* (NCU09564) and *pho-5* (NCU08325), as well as a glycerol diester phosphodiesterase-1 (NCU10038) whose main role is cleaving glycerol phosphate from cell membrane phospholipids. These genes are the major contributors to phosphate acquisition and are among the most highly differentially expressed genes across these two conditions (Figure 3-19). These specific phosphate acquisition genes are highlighted in green along with *nuc-1* and *nuc-2*. An additional large cluster of genes display high expression levels but are down-regulated under phosphate starvation. These genes are all ribosomal protein genes, indicating that *N. crassa* cells reduce expression of ribosomal proteins in response to the absence of phosphate.

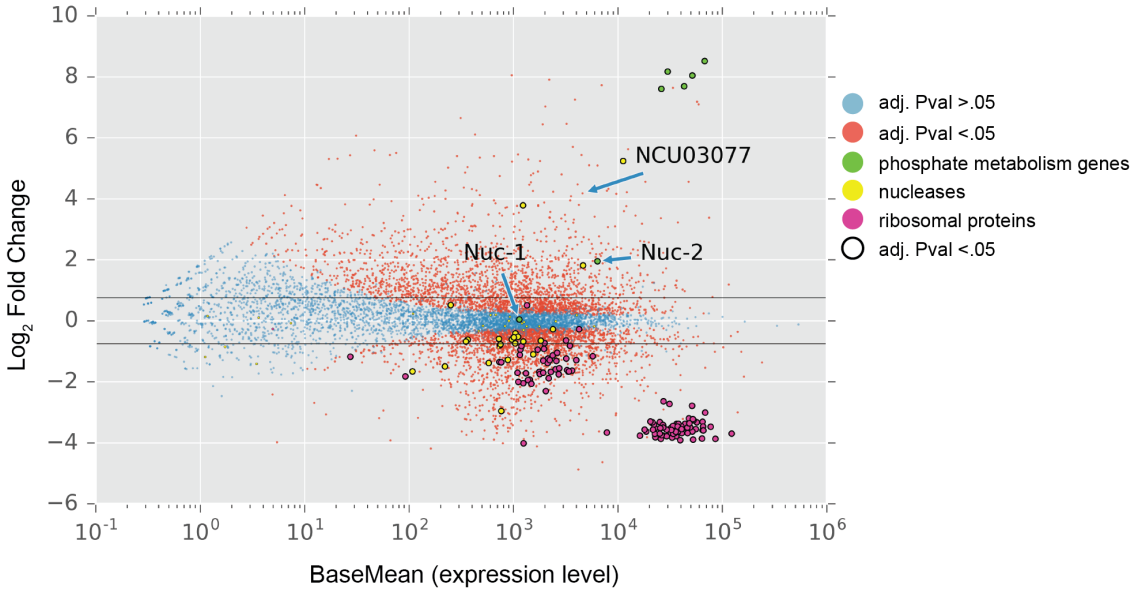


Figure 3-19 Differential expression analysis: High phosphate vs no phosphate. *Differential expression analysis between high KH_2PO_4(1g/L) and phosphate starvation conditions, where high KH_2PO_4 is the reference condition. Average expression of each gene is plotted along the x-axis while log_2 fold change is plotted on Y. Significantly differentially expressed genes with P adjusted values below 0.05 are represented in red, whereas statistically insignificant genes are represented in blue. Additional colors denote special groups of genes of interest, and outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.05)*

From this data we identified a candidate TF that might be involved with phosphate metabolism. We noticed that a single transcription factor (NCU03077) was up regulated specifically under phosphate starvation (Figure 3-19). This TF will be further discussed in section 4.6 along with a further exploration of the NUC-1 regulon and the connection between the two TFs.

3.6.2. *Response to sulfur starvation.*

Introduction

Like phosphate acquisition, *N. crassa* has a well-studied mechanism for sulfur acquisition. The fungus has a set of unlinked structural genes that are expressed only under sulfur-limited conditions. These include but are not limited to aryl sulfatase (NCU06041), choline sulfatase (NCU08364), high affinity methionine permeases (NCU04942, NCU07754), sulfate permease II (NCU04433) and a protease. These genes are regulated by a positive regulator, CYS-3 (NCU03536) (112, 142, 143), and a negative regulatory complex dubbed the SCF (Skp1p/Cullin/F-box) complex. It is made of several proteins encoded by SCON-2, SCON-3, and RBX-1 and several cullin encoding genes. This complex not only sequesters activity of CYS-3, it acts as E3 ubiquitin ligase designating proteins for proteolysis (144–146). When the *N. crassa* genome was sequenced, additional sulfur metabolic structural genes were identified and classified (147), however they have yet to be linked to CYS-3 and SCF regulation. We aimed to add an additional expression data to the genomic data and expand the understanding of the sulfur starvation response.

Results

N. crassa responds dramatically to sulfur limitation

Similar to our phosphate conditions, we profiled transcription under no, low and high sulfur conditions. Media was made by addition of methionine in a Vogel's media where magnesium sulfate was replaced with magnesium chloride so that methionine was the sole source of sulfur (see 3.2.1). When we compared these conditions we observed that few genes were differentially expressed between high sulfur and low sulfur conditions (data not shown). These results indicated to us that our low sulfur conditions still contained sufficient methionine after 4 hours for normal growth. However when we compared high sulfur to no sulfur or low sulfur to no sulfur, a far larger set of genes was differentially expressed. Our differential expression analysis between high sulfur vs no sulfur (high sulfur as reference condition), show a subset of sulfur acquisition, sulfur transport and sulfur regulation genes were up regulated (Figure 3-20). These genes are a complete set of sulfur acquisition and regulatory genes described in Borkovich et al. 2004 (147). We explore whether these genes are under positive regulation by CYS-3, and the full regulon of CYS-3 in chapter 5.

We further observed that, similar to phosphate starvation, sulfur starvation causes a great reduction of ribosomal protein expression. Ribosomal protein genes are denoted in pink in the scatter plot. We observe a cluster of these genes with high average expression as well as substantially reduced expression under sulfur starvation. Under sulfur starvation we also noticed that a number of PCWDEs (orange) and MFS transporters (purple) are slightly up regulated, indicating that sulfur starvation de-represses some of these genes.

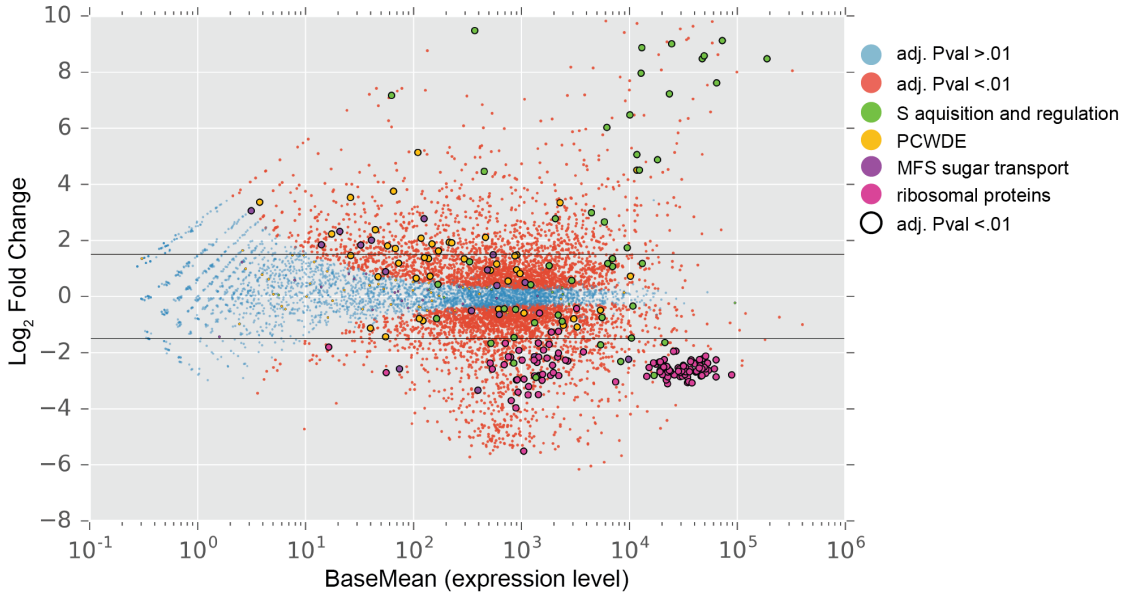


Figure 3-20 Differential expression analysis: high sulfur vs no sulfur. *Differential expression analysis between high sulfur (5mM methionine) and sulfur starvation (0mM methionine) conditions, where high sulfur is the reference strain. Average expression of each gene is plotted along the x axis while log_2 fold change is plotted on Y. Significantly differentially expressed genes with P adjusted values below 0.05 are represented in red, whereas statistically insignificant genes are represented in blue. Additional colors denote special groups of genes of interest, and outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.01)*

3.6.3. *N. crassa response to nitrogen starvation*

Introduction

Neurospora is able to utilize a diversity of nitrogen sources. Preferred nitrogen sources are ammonium, glutamine and glutamate. When these sources are unavailable, *N. crassa* can utilize many different nitrogen sources including nitrate, nitrite, purines, amides, most amino acids and proteins (51). Four transcription factors have been discovered that are involved in utilization of preferred and non-preferred nitrogen sources. These are *nit-2, nit-4, nmr* and *pco-1.* NIT-2, NIT-4 and NMR have been studied extensively regarding how they interact to regulate expression of nitrate reductase (*nit-3*), which is the critical enzyme for utilization of nitrate (51, 147). PCO-1 is involved in purine catabolism (148).

Nitrate is a non-preferred carbon source. Its conversion pathway to ammonium and glutamate/glutamine are shown in Figure 3-21 below.

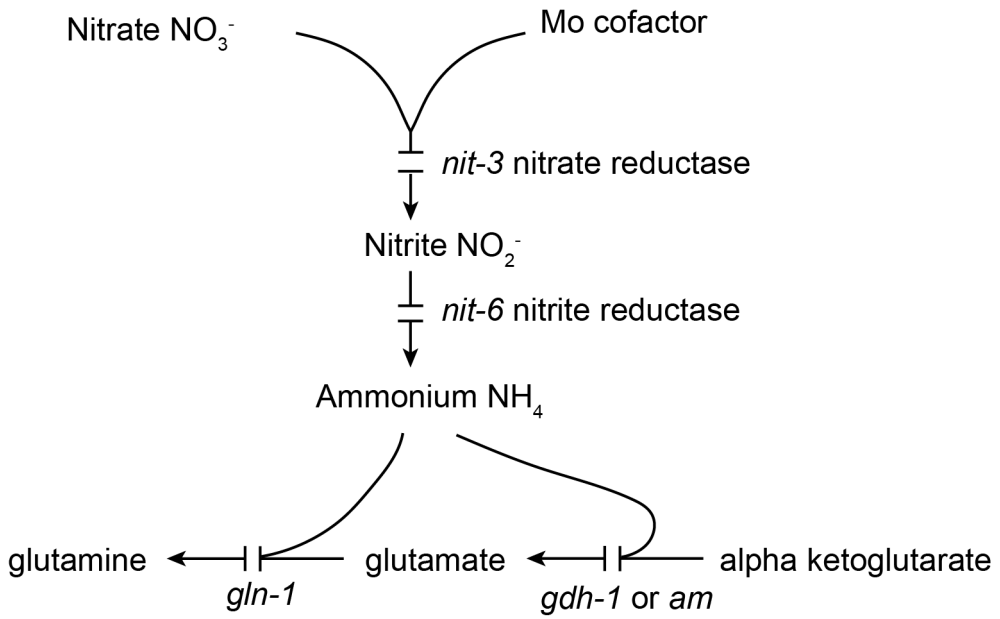


Figure 3-21 Pathway of nitrate to ammonium/glutamine *This figure is adapted from information presented in Marzluf et al. 1997* (53).

Gene products of *nit-3* and *nit-6* need to be present in order to convert nitrate to nitrite to ammonium. Much of the work exploring regulation of this pathway has focused on the *nit-3* gene encoding nitrate reductase. Three TFs coordinate expression of *nit-3.* Figure 3-22 summarizes information from a number of studies. The three transcription factors work cooperatively to only allow *nit-3* expression under conditions where preferred nitrogen sources are absent but nitrate is also present.

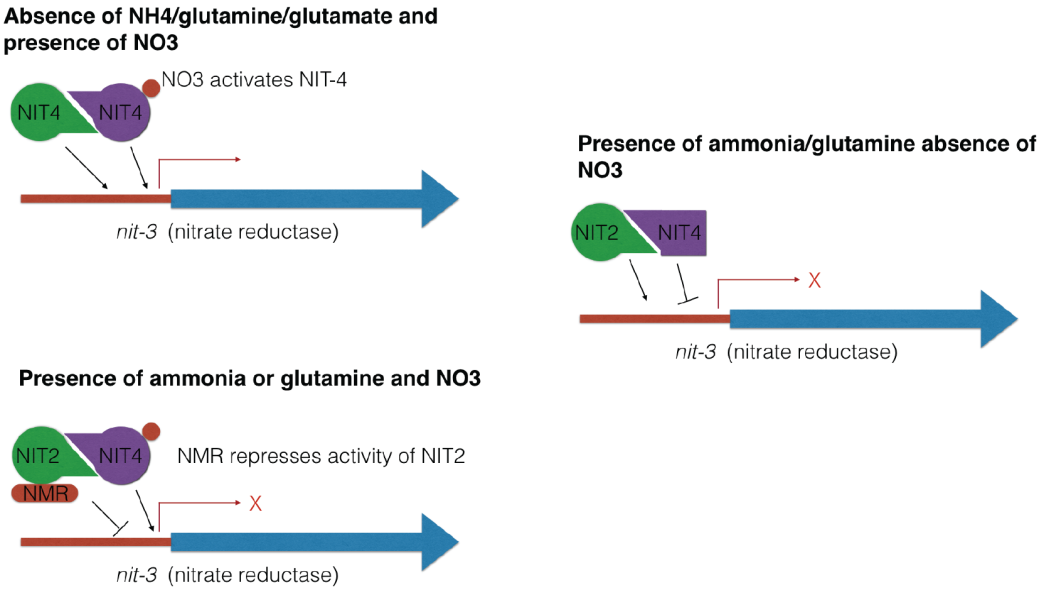


Figure 3-22 NIT-2, NIT-4, NMR regulation of *nit-3* expression. *This figure summarizes work from a number of studies* (149–151). *Nit-3 expression requires both active NIT-2 and NIT-4. In the presence of nitrate and absence of a preferred nitrogen source, NIT-4 is active along with active NIT-2 to promote expression of nit-3. In the presence of both preferred nitrogen source and nitrate, NIT-4 is activated by nitrate but NMR binds and represses activity of NIT-2 thus nit-3 is not expressed. In the presence of preferred nitrogen source and absence of nitrate, NIT-4 is inactive thus nit-3 is not expressed.*

We investigated the transcriptomes from conditions of preferred and non-preferred nitrogen sources, based on the metabolic models presented above. We used RNAseq and differential expression analysis conducted using DEseq to approach a more complete understanding of nitrogen metabolism. We found that conditional expression of *nit-3* and *nit-6* matches what was has been previously reported. We also find connections between nitrogen metabolism and phosphate, sulfur and carbon metabolism.

N. crassa responds dramatically to nitrogen starvation

We tested the response to nitrogen starvation in comparison to ammonium as the nitrogen source. The response was similar to responses observed in no sulfur and no phosphate. Genome wide expression changes take place along with reduced expression of ribosomal proteins and up regulation of nitrogen acquisition genes (Figure 3-23). This is not surprising as without nitrogen present, the cell is unable to continue normal cellular activities. The majority of the top differentially expressed genes are actually hypothetical proteins. This exemplifies how much is

still unknown regarding nitrogen starvation response. Within this list of top 50 up regulated genes include nitrogen acquisition genes including oligopeptide transporter 2 (NCU07894), amino acid permease inda1 (NCU07129), L-asparaginase (NCU05624), copper amine oxidase (NCU09406) and carboxypeptidase S1 (NCU05980). Unlike, phosphate and sulfur starvation, we do not observe a clear set of genes that are involved in nitrogen acquisition. Perhaps this can be explained by the diversity of nitrogen sources that can be utilized by *N. crassa.* Rather than over-expressing a small set of genes to acquire nitrogen, *N. crassa* may be employing a strategy utilizing moderate increased expression of a large number of genes involved in acquiring different sources of nitrogen. Thus no clear set of nitrogen acquisition genes can be distinguished from the differential expression data. Indeed, if we highlight secreted proteases, we see that the majority are not part of the top up regulated genes in nitrogen starvation. Additionally we find a small number of MFS sugar transporters and PCWDEs up regulated in nitrogen starvation; an indication that the regulation of nitrogen acquisition overlaps with carbon acquisition.

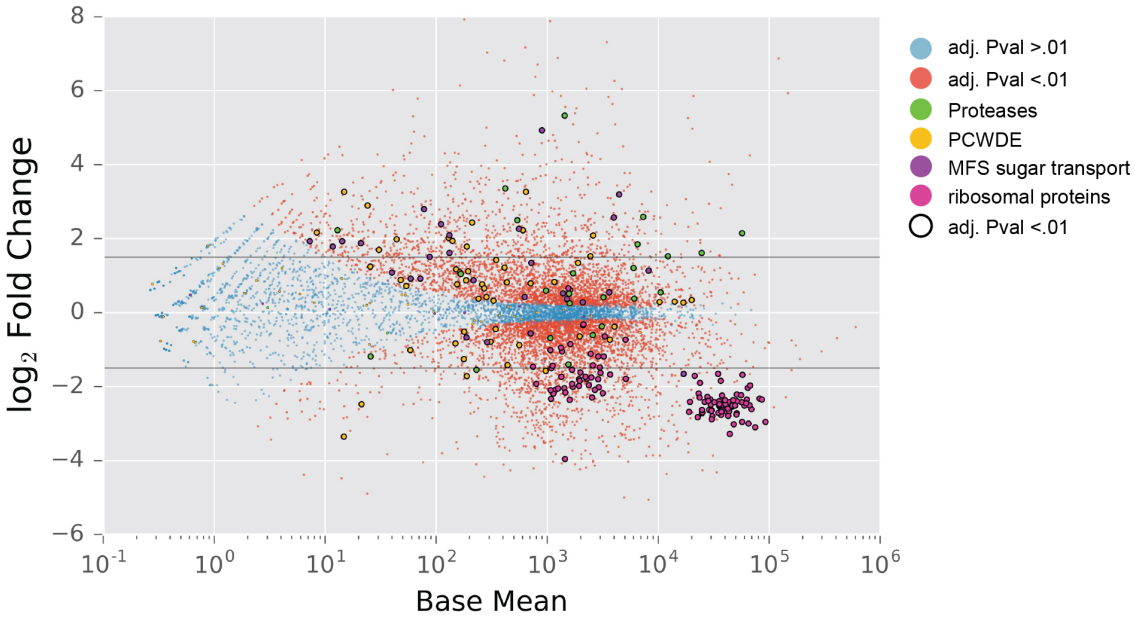


Figure 3-23 Differential expression analysis: WT ammonium vs no nitrogen
Differential expression analysis between ammonium and nitrogen starvation conditions, where ammonium is the reference condition. Average expression of each gene is plotted along the x axis while log_2 fold change is plotted on Y. Significantly differentially expressed genes with P adjusted values below 0.01 are represented in red, whereas statistically insignificant genes are represented in blue. Additional colors denote special groups of genes of interest, and outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.01)

Presence of nitrate and absence of preferred carbon source recapitulates previous research and reveals crosstalk between N, C, P, S metabolism

Ammonium versus nitrate differential expression analyses support *nit-3* canon, but also reveal additional genes that may be regulated by this process. Under nitrate conditions, we expect to see the up regulation of the structural genes required for nitrate utilization in the absence of a preferred nitrogen source like ammonium. These structural genes include *nit-3* nitrate reductase, *nit-6* nitrite reductase, *CRNA* nitrate permease. As expected, these genes are the top 3 differentially expressed genes in our ammonium vs nitrate dataset. The fourth most differentially expressed genes between these two conditions is NCU10051 flavohemoglobin. Although we do not know the function of this gene, we hypothesize that this gene also requires NIT-4 and NIT-2 in order to be expressed. Surprisingly, within the top 20 genes differentially expressed between these two conditions, we see several crucial phosphate and sulfur acquisition genes up regulated on nitrate. These include inorganic phosphate transporter *pho-5* (NCU08325), phosphate permease *pho-4* (NCU9564), alkaline phosphatase *pho-2* (NCU01376), aryl-sulfatase-1 (NCU06041), and two taurine dioxygenases (NCU01057 and NCU07610). This is a clear indication that nitrogen metabolism is intricately connected with sulfur and phosphate acquisition. Importantly though, these genes are far from their expression levels associated with sulfur and phosphate limiting conditions respectively. Furthermore, we observe in our plot that a handful of PCWDEs and MFS sugar transporters are also up regulated on nitrate, indicating cross talk between nitrogen and carbon metabolism.

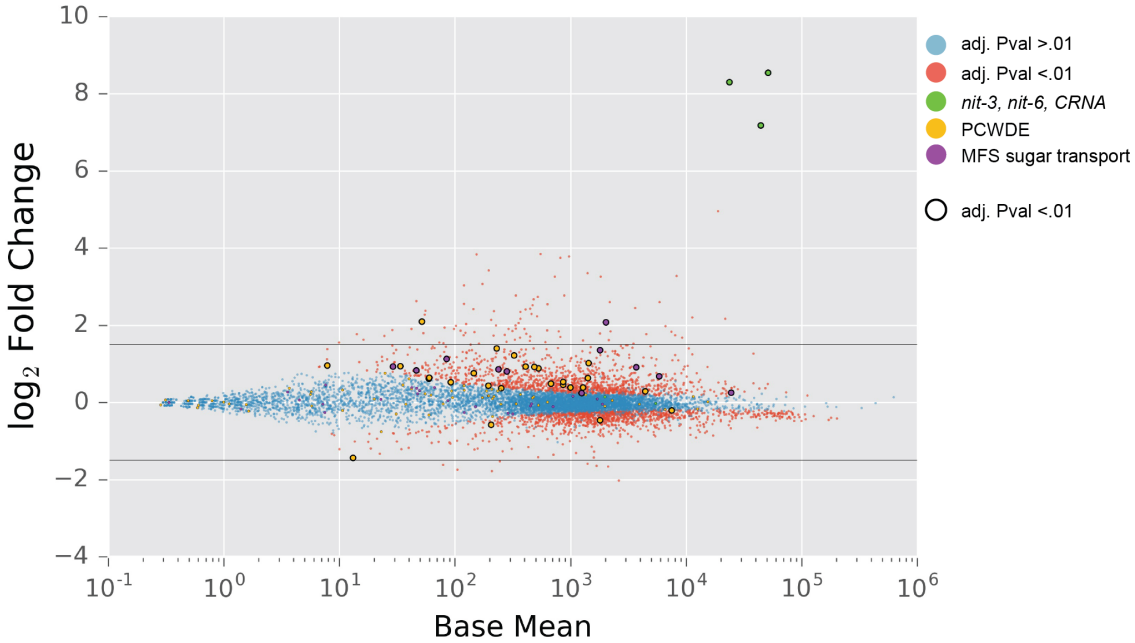


Figure 3-24 Differential Expression analysis: ammonium vs nitrate *Differential expression analysis between ammonium and nitrate conditions, where ammonium is the reference condition. Average expression of each gene is plotted along the x axis while log_2 fold change is plotted on Y. Significantly differentially expressed genes with P adjusted values below 0.01 are represented in red, whereas statistically insignificant genes are represented in blue. Additional colors denote special groups of genes of*

interest, and outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.01)

Our nitrogen dataset contains a much wider range of nitrogen conditions that we have not further explored. These include conditions where specific amino acids are the sole nitrogen source. We believe much more can be learned form this dataset regarding metabolism of different nitrogen sources.

3.7. Network analysis of full RNAseq dataset reveals sets of co-regulated genes

3.7.1. *Introduction*

This section focuses on analyses conducted by Sara Calhoun at the Joint Genome Institute. The analysis presented here was conducted by Sara using the RNAseq data sets generated from this project.

Sara used weighted gene co-expression network analysis (WGNCA) (152) to determine "modules" or sets of genes that are co-regulated together across all of our RNAseq data. The basic idea behind this method involves constructing a network where genes are connected to one another based on an adjacency metric derived from correlation of expression profiles. Genes that do not show changes in expression or expressed at very low levels are filtered out. Genes that have expression profiles that do not conform to any module are also filtered out. By using expression profiles of 2392 genes, Sara ended up with a network of modules that also show varying degrees of connectedness to other modules (Figure 3-25).

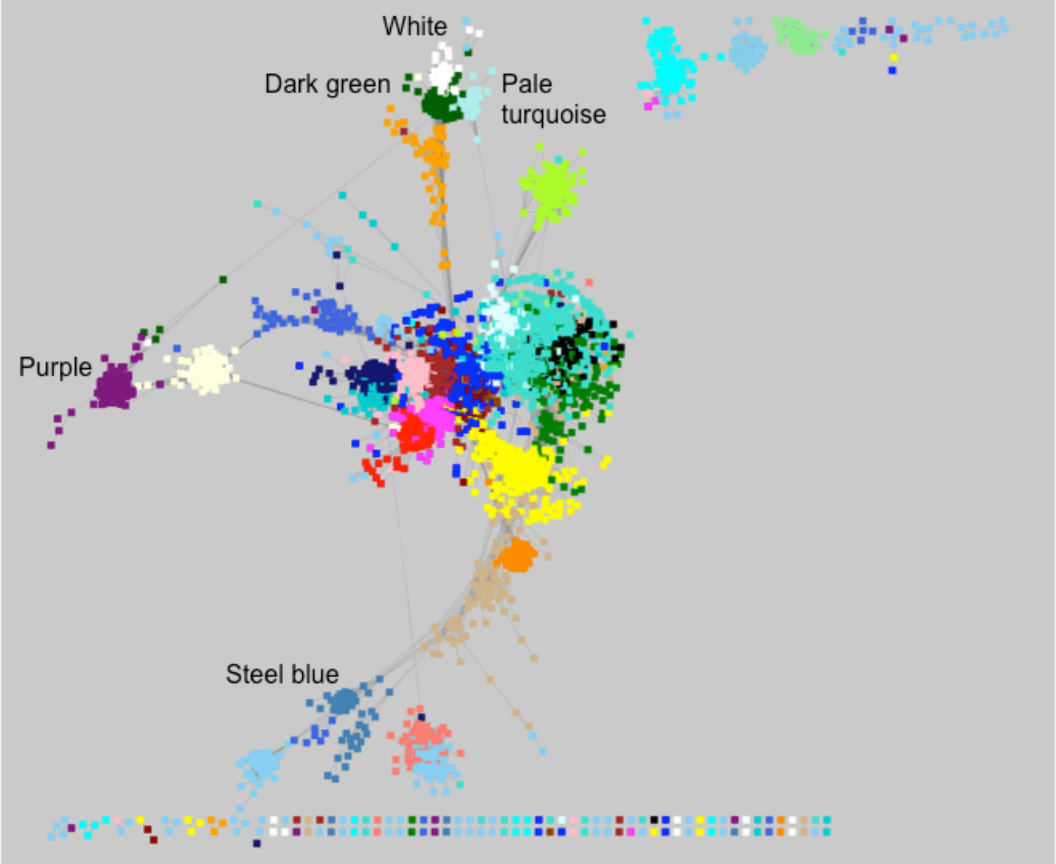


Figure 3-25 Weighted Gene Co-expression Network Analysis using full RNAseq dataset. *Color sets or "Modules" that are labeled are enriched for of cellulose, hemicellulose and pectin catabolic genes. These include Purple, Dark green, White, Pale turquoise and Steel blue modules.*

3.7.2. ***Exploring WGCNA modules***

Sara further explored the genes in each modules using FUNCAT (121) with a updated database described in Thieme et al. 2017 (35). Many of modules are not functionally enriched for any processes in particular, and contain many unclassified genes. However, other modules are enriched for notable processes and functions. The modules labeled with color name are enriched for various processes within carbon metabolism. Below are some of the carbon metabolism related modules, along with a few other modules of interest. Included in each module are the top 10 functional categories. This provides some information into the processes a module's genes are involved in.

purple

01.05.03 polysaccharide metabolism	4.29E-09
01.05.03.06.07 hemicellulose/pectin catabolism	9.33E-08
01.05.03.06 hemicellulose/pectin metabolism	9.33E-08
01.25 extracellular metabolism	4.39E-07
01.25.01.03 extracellular hemicellulose/pectin	1.15E-06
14.13.04.02 vacuolar protein degradation	4.08E-06
01.05.03.06.07.02 specific pectin catabolism	4.21E-06
01.05.03.06.07.01 specific hemicellulose catabolism	9.06E-06
14.13.04.01 lysosomal protein degradation	1.03E-05
01 METABOLISM	1.32E-05

pale turquoise

01.25.01.02 extracellular cellulose degradation	1.07E-13
32.10.07 degradation / modification of foreign	4.11E-13
16.05 polysaccharide binding	5.15E-13
01.25.07 extracellular ester compound degradation	5.76E-13
01.05.03.05 cellulose metabolism	2.92E-12
01.05.03.05.07 cellulose catabolism	2.92E-12
01.25.01 extracellular polysaccharide degradation	4.94E-12
01.25 extracellular metabolism	5.58E-12
01.25.01.03.01 extracellular hemicellulose specific	3.11E-10
32.10 degradation / modification of foreign	6.77E-10

white

01.05 C-compound and carbohydrate metabolism	1.00E-14
01.20.19.05 metabolism of cobalamins	2.83E-12
01.05.03.06 hemicellulose/pectin metabolism	1.25E-11
01.05.03.06.07 hemicellulose/pectin catabolism	1.25E-11
01.05.03.06.07.01.01 xylan catabolism	4.11E-09
01.20.38 metabolism of toxins/drugs	4.82E-09
01.20.19 metabolism of secondary products derived	1.27E-08
01.05.03.06.07.01 specific hemicellulose catabolism	1.46E-08
01.05.03 polysaccharide metabolism	3.11E-08
32.07.03 detoxification by modification	3.22E-08

Darkgreen

01.25.01 extracellular polysaccharide degradation	1.70E-33
01.25 extracellular metabolism	7.49E-31
01.25.01.03.01 extracellular hemicellulose specific	5.52E-30
01.25.01.03 extracellular hemicellulose/pectin	4.91E-29
01.25.01.03.01.01 extracellular xylan degradation	8.60E-29
01.05.03.05 cellulose metabolism	1.76E-28
01.05.03.05.07 cellulose catabolism	1.76E-28
01.05.03.06.07 hemicellulose/pectin catabolism	7.88E-28
01.05.03.06 hemicellulose/pectin metabolism	7.88E-28
01.25.07 extracellular ester compound degradation	1.66E-27

steel blue

01.05.03.06.07 hemicellulose/pectin catabolism	1.88E-11
01.05.03.06 hemicellulose/pectin metabolism	1.88E-11
01.05.03.06.07.02.01 homogalacturonan catabolism	5.31E-09
01.05.03.06.07.02 specific pectin catabolism	9.50E-09
01.25.01.03 extracellular hemicellulose/pectin	1.41E-07
01.25 extracellular metabolism	1.76E-07
01.05 C-compound and carbohydrate metabolism	5.56E-07
01.25.01.03.02.01 extracellular homogalacturonan	5.81E-07
01.05.03 polysaccharide metabolism	7.16E-07
01.25.01.03.02 extracellular pectin specific degradation	2.89E-06

brown

11.04.01 rRNA processing	2.88E-72
11.04 RNA processing	6.97E-65
11 TRANSCRIPTION	3.68E-51
12 PROTEIN SYNTHESIS	6.29E-50
16.03.03 RNA binding	1.90E-44
12.01 ribosome biogenesis	5.40E-43
11.02.01 rRNA synthesis	1.03E-42
16.03 nucleic acid binding	5.53E-37
16 PROTEIN WITH BINDING FUNCTION OR	9.02E-33
11.06 RNA modification	1.87E-28

black

16.17.05 sodium binding	0.000297057
20.09.16 cellular export and secretion	0.000332746
20.09.07.05 intra Golgi transport	0.001701779
20.01.26 neurotransmitter transport	0.00597931
34.07.01 cell-cell adhesion	0.006472415
42.02 eukaryotic plasma membrane	0.006917004
99 UNCLASSIFIED PROTEINS	0.006970417
42.08 Golgi	0.007875015
18.02.01.02.01 GTPase inhibitor (GIP)	0.010685017
20.09.07.06 post Golgi transport	0.011549141

orange

14 PROTEIN FATE (folding, modification, destination)	1.87E-21
20.09.07 vesicular transport (Golgi network, etc.)	4.38E-21
14.04 protein targeting, sorting and translocation	5.00E-21
20.01.10 protein transport	6.08E-21
20.09.07.03 ER to Golgi transport	3.91E-20
20.09.05 non-vesicular ER transport	3.36E-18
20.09 transport routes	9.18E-18
20 CELLULAR TRANSPORT, TRANSPORT FACILITIES	1.26E-13
14.07.02 modification with sugar residues (e.g.	1.04E-12
16.01 protein binding	1.23E-12

Figure 3-26 Key carbon WGCNA modules *Each module contains the top 10 enriched Funcat processes, along with the p value for their enrichment.*

The Purple, Pale turquoise, White, Dark green, and Steel blue modules are the modules that are enriched for PCWDEs. These modules have a mixture of PCWDEs degrading different polysaccharides. The distance between these different modules along with their heterogeneity of function indicate that PCWDEs are not all co-regulated by their function. These data indicate that PCWDEs capable of degrading different polymers can be co-expressed and are thus likely to be co-regulated. This observation has already been noted in the past, where TFs like CLR-2 and XLR-1 have been shown to regulate more PCWDEs than just cellulases and xylanases, respectively. According to our modules, pectinases may behave similarly, as they are grouped in the same modules that many cellulases and hemicellulases.

The Brown and Orange modules represent informative non-carbon metabolic modules. The Brown module is enriched for RNA processing and synthesis genes. It is no surprise that these genes form a module. Transcription requires all of these processes. The Orange module is one of the most interesting modules in relation to polysaccharide degradation. This module is composed of endoplasmic reticulum related processes, such as protein folding and modification of sugar moieties. This module also clusters closely with the White, Pale Turquoise and Dark Green modules. These data indicate that genes in the Orange module may be important for the increase of protein production associated with cellulase and hemicellulase production.

The Black module is an example of a less informative module. Enrichment p values are relatively high, and the categories that are enriched are unrelated to one another. The majority of the modules fit this description. The Black module in particular contains categories of cell-cell adhesion, sodium binding, post Golgi transport and others, all seemingly unrelated to each other. The observation that these genes are co-expressed indicates that their remains much to be discovered regarding the processes that take place in the fungal cell.

3.7.3. ***Discussion***

In summary, WGCNA analysis informed us of a several interesting things. PCWDEs do not form modules according to function, indicating that their regulation is more complex than a scheme where a single transcription factor regulates the catabolism of a single polysaccharide. Modules also provide candidate genes for further investigation. For instance, genes within carbon modules that are not PCWDEs may still be important factors for degrading and utilizing the plant cell wall. Genes in the Orange module, which closely cluster with carbon modules, may be important for secretion of PCWDEs. Finally, we find the majority of the modules uninformative; an indication that there is still much to understand regarding fungal cell processes.

Chapter 4. **Transcriptional regulation of carbon metabolism**

4.1. Introduction

This chapter focuses on characterizing our candidate transcription factors associated with plant cell wall degradation. From the candidate list we generated, we induced each ΔTF strain with the conditions where it showed the highest expression. We conducted RNAseq to determine what genes might be affected by the deletion of each TF and compared that data back to WT data on the same condition. Table 4-1 summarizes the deletion strains we tested and the conditions that we tested them under.

Deletion	condition	Deletion	condition
NCU00282	2mM_fructose	NCU04058	2mM_ribose
NCU00289	1%_amylose	NCU04158	No nitrogen
NCU00445	2mM_mannose	NCU04158	2%_sucrose
NCU00445	1%_glucomannan	NCU04211	2mM_mannose
NCU00808	1%_galactan	NCU04295	no_sulfur
NCU00808	1%_citrus_peel	NCU04295	1%_citrus_peel
NCU00808	1%_arabinan	NCU04848	1%_switchgrass
NCU00890	2mM_mannobiose	NCU04851	2mM_rhamnose
NCU01008	1%_energy_cane	NCU05024	1%_galacturonic_acid
NCU01074	1%_xylan	NCU05414	1%_miscanthus
NCU01154	1%_pectin	NCU05909	1%_xylan
NCU01154	1%_citrus_peel	NCU05909	2mM_xylose
NCU01209	1%_mixed-linkage_glucan	NCU06173	1%_mixed-linkage_glucan
NCU01312	1%_miscanthus	NCU06173	1%_Avicel
NCU01312	2mM_fucose	NCU06920	1%_galacturonic_acid
NCU01386	1%_energy_cane	NCU07728	1%_mixed-linkage_glucan
NCU01640	1%_amylose	NCU08055	2mM_ribose
NCU02203	2mM_ribose	NCU08634	1%_amylose
NCU02307	no_carbon	NCU08899	2mM_rhamnose
NCU02853	1%_amylose	NCU09169	1%_citrus_peel
NCU03077	no_phosphate	NCU09169	1%_pectin
NCU03421	1%_miscanthus	NCU09252	1%_amylose
NCU03417	2mM_ribose	NCU09529	no_sulfur
NCU03643	2mM_rhamnose	NCU10697	1%_citrus_peel
NCU03699	1%_Avicel		

Table 4-1 List of ΔTFs and conditions tested

Unfortunately, many TFs deletion mutants tested showed no statistically significant change in gene expression between the ΔTF and WT strain under the conditions tested. We conducted no further analysis for these TFs. There were also TFs that showed changes in gene expression but no cohesive function could be concluded from the differentially expressed gene set. We categorize some of the TFs

in this section, and propose some hypothesis for the processes they might be involved in. We also found TFs involved in carbon metabolism. For minor TFs, we explore the information learned from RNAseq. For more important TFs, we explore their function further using genetic and biochemical experiments. This section is divided into these three different categories of TFs, with an additional section regarding a phosphate acquisition TF that did not fit elsewhere in the dissertation.

The major transcription factors that we identified were all involved in hemicellulose and pectin degradation. Specifically, they are important for pectin backbone degradation and arabinan/arabinose metabolism. Unlike cellulose and xylan, the metabolism of these substrates has remained mostly unexplored in *N. crassa.* We explore these TFs in the context of the metabolism of these pectin as well as arabinan/arabinose in order to fully appreciate their regulatory role. Additionally, we found single TFs systems that act independently but have highly overlapping regulons.

4.2. Characterizing the role of NCU09033 (PDR-1) and NCU04295 (PDR-2) in pectin metabolism

4.2.1. *Introduction*

Pectin is an abundant plant polymer in the primary cell wall. This cell wall type associated with newly growing tissues as well as fruit and leaf tissue(153). Pectin encompasses a number of complex polysaccharides that are rich in 1,4-linked α-D-galacturonic acid (GalA), but also contain a high diversity of other sugars. The two predominant polymers that constitute pectin are homogalacturonan (HG) and rhamnogalacturonan 1 (RG1). HG has a unsubstituted backbone of GalA, while RG1 has alternating GalA and rhamnose residues, with side chains that are composed of mainly arabinose and galactose (17). A third less abundant domain of pectin named rhamnogalacturonan II (RG2) is the most complex of the described pectin components and is composed of 12 different monosaccharides (154). Additional sugars that are found in pectin include xylose, glucuronic acid, fucose and others (155).

The metabolic pathway of GalA has been studied in several other ascomycete species, but the pathway has not been biochemically confirmed in *N. crassa*. In *T. reesei*, GalA metabolism begins with four key reactions. GalA is converted to L-galactonate, to 2-keto-3-deoxy-L-galactonate, then split into pyruvate and L-glyceraldehyde (156, 157). L-glyceraldehyde is then converted to glycerol (157), which can be further reduced and enter into glycolysis. The enzyme that catalyzes the first step in the pathway, galacturonic acid reductase (GaaA), has been described in *T. reesei* (156); an ortholog is present *N. crassa* (NCU01906). Interestingly another GaaA orthologous to *A. niger* GaaA (An02g07720) (158) is present in *N. crassa* (NCU09533). The second enzyme in the pathway, L-galactonate-dehydratase (GaaB)(159), has a single homolog (NCU07064) and which is under the control of the pectinolytic regulator PDR-1 (35). The third enzyme in the pathway, L-threo-3-deoxy-hexulosonate aldolase (GaaC), has also been biochemically verified in *T. reesei* (160), and its ortholog is present in *N. crassa* (NCU09532), adjacent to the GaaA

NCU09533. The 4th enzyme in the pathway, an L-glyceraldehyde reductase (GaaD) has been putatively described in *A. niger* (158), but confusingly its closest homolog in *N. crassa* is the same GaaA that was found in *T. reesei* (NCU01906). No biochemical verification has been conducted in *A. niger.* BLAST (161) results show two additional homologs, NCU04923 and NCU04510, that are candidates for GaaD and have ~40% identity with the *A. niger* enzyme. Both are annotated as being part of glycerol metabolism pathway. We assume that NCU01906 is the true GaaD, as its transcription profile is most specific for pectinaceous substrates (data not shown). We conclude that a complete GalA metabolic pathway is present in *N. crassa*, but there is still confusion regarding which enzymes catalyze each step.

The metabolic pathway for rhamnose has also been described in fungi, and is likely conserved in *N. crassa.* The pathway is different from the bacterial pathway and was first described in the yeast-like species *Aureobasidium pullulans* (162) and then confirmed in the yeast species *Pichia stipitis* and *Debaryomyces polymorphus* (163). In this pathway, rhamnose is converted to L-rhamnono-γ-lactone to L-rhamnonate, to L-2-keto-3-deoxyrhamnonate (L-KDR) and finally to pyruvate and L-lactaldehyde. The enzymes that catalyze these reactions are L-rhamnose-1-dehydrogenase (LRA1), L-rhamnono-γ-lactonase (LRA2), L-rhamnonate dehydratase (LRA3), and L-KDR aldolase (LRA4). *N. crassa* has homologs to each of these enzymes: NCU09035, NCU3605, NCU09034 and NCU05037, respectively. LRA4 has a second homolog as well (NCU05977) with 30% identity to the *P. stipitis* LRA4, however, of the two homologs only NCU05037 is specifically up regulated on rhamnose (Figure 4-1).

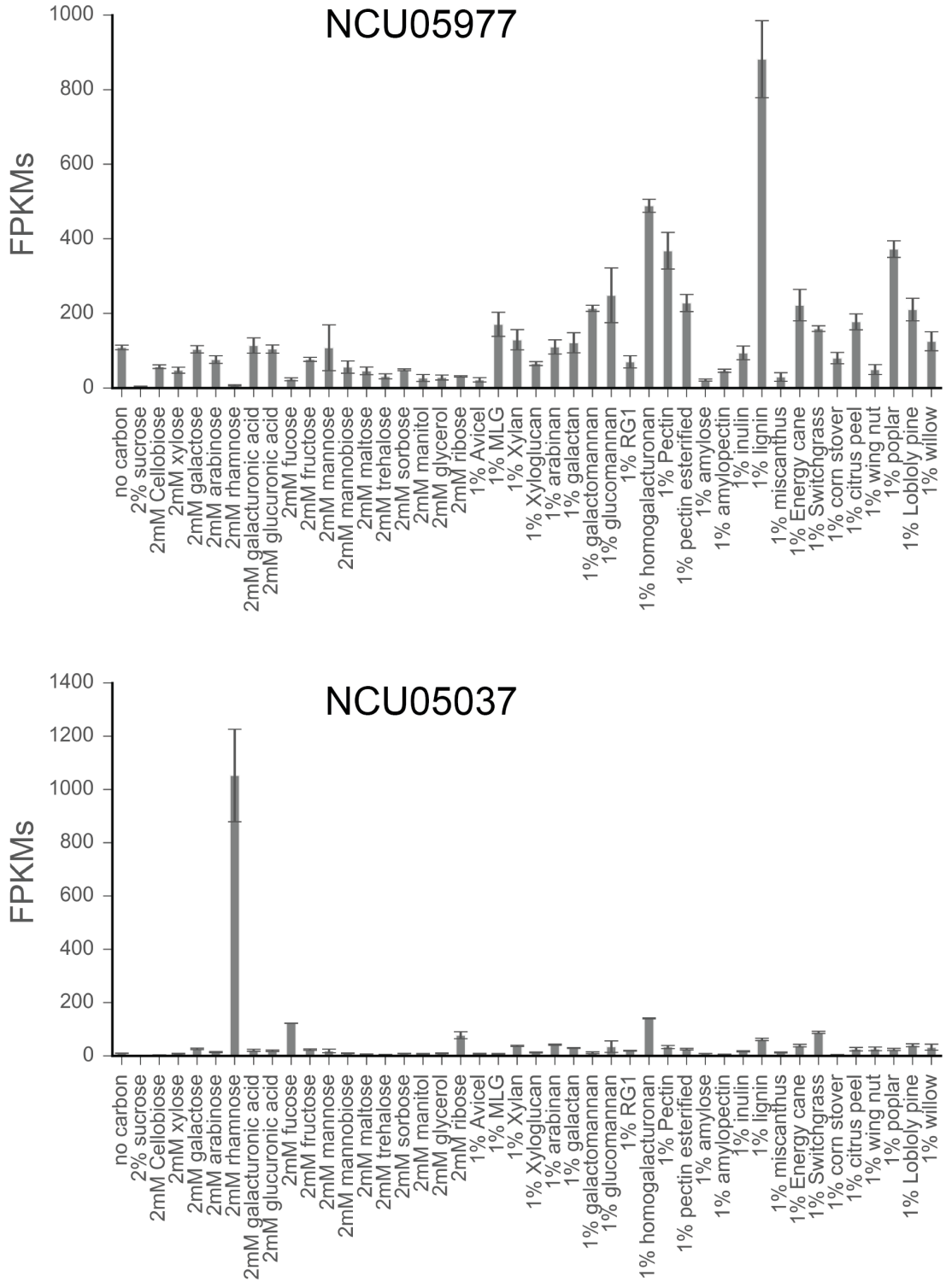


Figure 4-1 Expression of putative LRA4s across WT conditions. Expression measured by FPKMs with error bars representing standard deviation from triplicate

N. crassa also has a suite of extracellular PCWD pectin-specific enzymes and sugar transporters present in the genome. The most important enzymes for pectin degradation are those that degrade the pectin backbone. For degrading HG, *Neurospora* has several pectate lyases and galacturonanases, and a pectin methyl esterase. The organism has several additional enzymes to degrade RG1 backbone, which include rhamnogalacturonan lyase, rhamnogalacturonan acetylesterase and a putative unsaturated rhamnoglacturonyl hydrolase. Further degradation of side chains carried out by a suite of enzymes, which include, xylosidases, arabinases, galactosidases, glucuronidases, feruloyl esterases and more. Many of these pectin side-chain degrading enzymes are also extremely important for degrading hemicellulose. Below is a table summarizing the enzymes present in the *N. crassa* genome specific for pectin metabolism (Table 4-2).

Gene ID	Annotation	process
NCU09533	Galacturonic acid reductase GaaA	GalA metabolism
NCU07064	L-galactonate-dehydratase GaaB	GalA metabolism
NCU09532	L-threo-3-deoxy-hexulosonate adolase GaaC	GalA metabolism
NCU01906	Galacturonic acid reductase (GaaA) or L-glyceraldehyde reductase (GaaD)	GalA metabolism
NCU00988	GalA transporter (Gat-1)	GalA metabolism
NCU09035	L-rhamnose-1-dehydrogenase LRA1	rhamnose catabolism
NCU03605	L-rhamnono-γ-lactonase LRA2	rhamnose catabolism
NCU09034	rhamnonate dehydratase LRA3	rhamnose catabolism
NCU05037	L- and L-KDR aldolase LRA4	rhamnose catabolism
NCU06326	pectate lyase 1	backbone degradation
NCU08176	pectate lyase A	backbone degradation
NCU10045	pectin methyl esterase	backbone degradation
NCU06961	exogalacturonanase	backbone degradation
NCU02369	endogalacturonanase	backbone degradation
NCU05598	rhamnogalacturonan lyase	backbone degradation
NCU09976	rhamnogalacturonan acetyl esterase	backbone degradation
NCU05852	glucuronan lyase A	sidechain degradation
NCU00937	extracellular β-glucuronidase	sidechain degradation
NCU09774	feruloyl esterase	sidechain degradation
NCU09491	feruloyl esterase	sidechain degradation
NCU08785	feruloyl esterase	sidechain degradation
NCU00972	endo-β-1,4-galactanase	sidechain degradation
NCU05882	endo-β-1,6-galactanase	sidechain degradation
NCU09702	endo-β-1,6-galactanase	sidechain degradation

Table 4-2 Genes involved in Pectin metabolism

We explore the regulons of two transcription factors, PDR-1 and PDR-2 in this section. These two transcription factors are responsible for positive regulation of the majority of the pectin metabolic genes presented above.

4.2.2. ***Results***

PDR-1 combines certain aspects of the roles of RhaR and GaaR and AraR

The majority of the work on PDR-1 was conducted by Nils Thieme, Axel Dietschmann and Philipp Benz at the Technical University of Munich. I contributed RNAseq experiments of WT and the Δ*pdr-1* mutant on several carbon sources, and some of data analysis regarding these experiments, and final edits of published manuscript (35). For this reason, I include only a short summary of our key findings in this section in order to have a more complete understanding of pectin metabolism in *N. crassa.*

Pdr-1 was originally found in a screen of transcription factor deletion mutants that had defective growth on pectin. Δ*pdr-1* had a severe growth defect specific for pectin and also showed reduced growth on pectin specific carbon sources. Most severe was growth on rhamnose and arabinose and polygalacturonic acid (PGA), and moderately on GalA. This indicated that PDR-1 was critical for rhamnose and arabinose metabolism and PGA depolymerization, as well as moderately important for GalA metabolism. RNAseq of WT compared to Δ*pdr-1* strain grown on 2mM rhamnose and 1% pectin showed that genes crucial for in rhamnose catabolism (LRA1, LRA3 Table 4-2), arabinose catabolism (LADH Table 4-3), GalA metabolism (GaaC, GaaD Table 4-2), and PGA depolymerization (NCU02369 endogalacturonanase) were all down regulated in the Δ *pdr-1* strain as compared to WT, suggesting that these genes require PDR-1 for full expression. This data can explain much of the growth phenotypes observed in Δ *pdr-1*. The growth phenotype of Δ *pdr-1* on pectin and PGA was also not solely due to down regulation of endogalacturonanase NCU02369, as overexpression of NCU02369 did not rescue Δ *pdr-1* growth. Finally, in an overexpression strain of *pdr-1*, endogalacturonanase activity only increased in the presence of rhamnose and not other sugars, indicating that PDR-1 requires rhamnose to be present in order positively regulate gene transcription. Rhamnose may somehow allow better translocation of the PDR-1 to the nucleus. The presence of rhamnose increases intensity of GFP in the nucleus of a PDR-1-GFP strain. In summary, *pdr-1* helps regulates rhamnose, arabinose and GalA metabolism, as well as polygalacturonic depolymerization, but is only active in the presence of rhamnose (35).

PDR-2 regulates GalA metabolism and the majority of extracellular pectinolytic enzymes

PDR-2 was one of the initial transcription factors candidates chosen for its transcriptional profile across our carbon conditions. Gene expression of *pdr-2* increased on 1% pectin, 1% citrus peel, but was reduced expression under 2%

sucrose conditions. Expression reduction on sucrose is characteristic of TFs that regulate PCWDEs (Figure 4-2). The majority of these TFs have reduced expression under sucrose conditions, likely due to carbon catabolite repression.

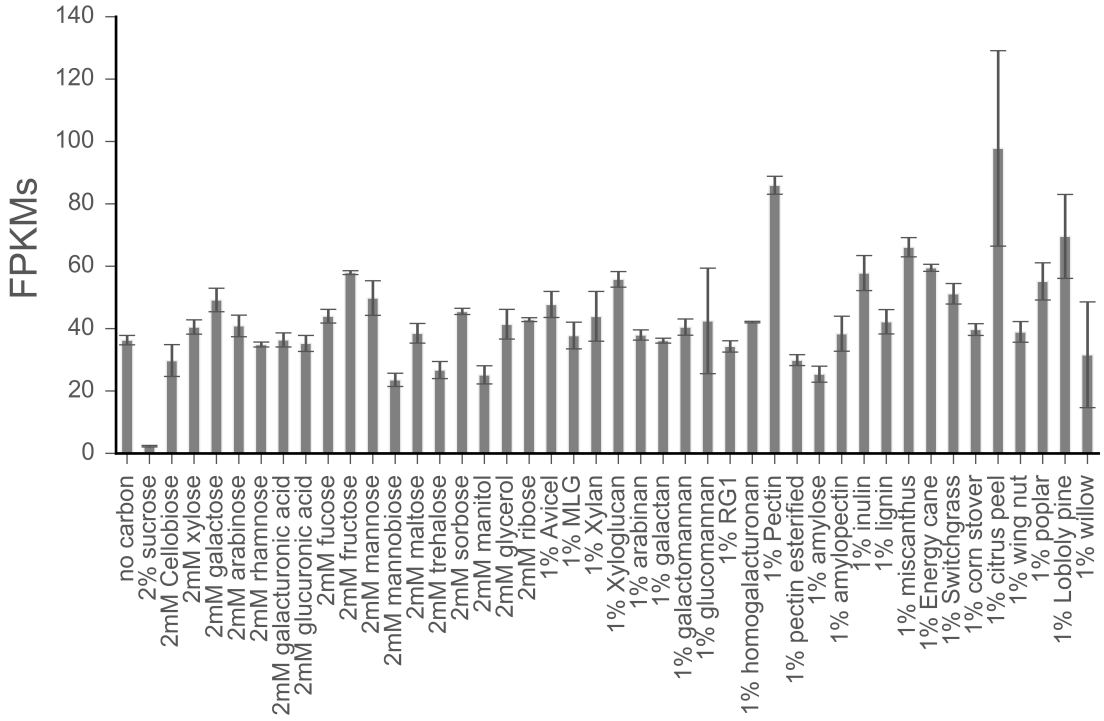


Figure 4-2 *pdr*-2 Expression Across Carbon Conditions *Expression as measured by FPKMs with error bars representing standard deviation from triplicate.*

RNAseq of WT and Δ*pdr-2* induced by 1% citrus peel shows that PDR-2 strongly regulates expression of a number of key galacturonic acid catabolism genes in addition the majority of pectinolytic genes present the *N. crassa* genome. To determine genes regulated by PDR-2 we conducted differential expression analysis between Δ*pdr-2* and WT using DEseq software (117). Genes differentially expressed positively, where fold change of expression between Δpdr-2 and WT is positive, are likely regulated by positively by PDR-2. When we plot the fold change for each gene across their average expression level (base mean), we see that PDR-2 positively regulates a number of genes (Figure 4-3A). Fold change values for each gene indicate the level of influence PDR-2 has on expression of that gene. When we only look at the fold change of the genes responsible for pectin catabolism, we see that that the top differentially expressed genes in these conditions are a exopolygalacturonanase, two pectate lyases, *gaaA, gaaB, gaaC, gat*-1 and a small number of arabinases (Figure 4-2B). This list includes three of the four major enzymes that act on homogalacturonan, which makes up the majority of the pectin backbone. These data indicate that Pdr-2 is important for cleavage of the pectin backbone, GalA transport and intracellular GalA catabolism. One major pectinase that is missing from this list is the polygalacturonanase NCU02369. Expression of

this gene, however, is under strong regulation by PDR-1. PDR-2 seems to exert moderate control over rhamnose catabolism through modulation of LRA1 and LRA3 as well, and minor control of arabinose metabolism through modulation of *lat-1* arabinose transporter and LADH. In conclusion, PDR-2 is a major regulator of polygalacturonic acid depolymerization and GalA catabolism, with a minor role in the catabolism of other pectinaceous sugars.

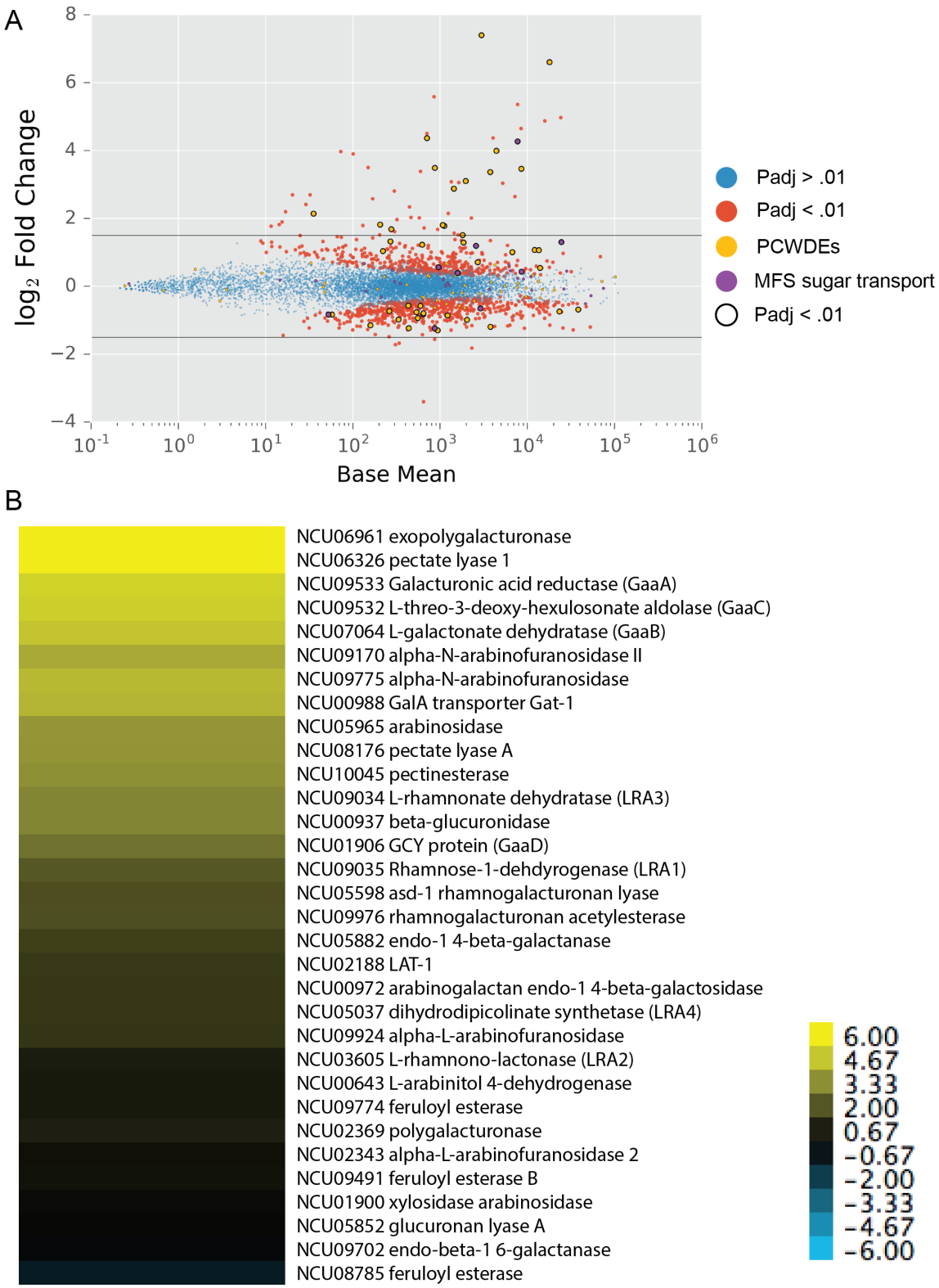


Figure 4-3 Differential Expression analysis of *Δpdr-2* vs WT on 1% citrus peel

A) Differential expression as data calculated by DEseq plotted by Log_2 fold change vs

base mean for Δpdr-2 vs WT on 1% citrus peel as the sole carbon source. Δpdr-2 is the reference condition. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj = 0.01. B) Heat map of differential expression of pectin catabolic genes as measured by log_2 fold change between ΔPdr-2 and WT. Color bar represent direct log_2 fold change values.

4.2.3. ***Discussion***

The regulation of pectin catabolism appears to be well conserved between *Aspergillus* and *Neurospora. RhaR* and *GaaR* genes encode for two transcription factors in *Aspergillus* and others that regulate expression of rhamnose and GalA metabolism, respectively. They also positively regulate extracellular PCWDEs responsible for cleaving these sugars from pectin polymers (36–38). *Pdr-1* and *pdr-2* are homologs of these two genes in *N. crassa*. They show conservation in function with *Aspergillus* homologs. The two TFs encoded by these genes act cooperatively regulate expression of genes required for pectin catabolism. Like RhaR and GaaR, each TF is activated by rhamnose and GalA respectively. The mechanisms for RhaR activation are unknown.

A repressor of GaaR has also been discovered in A*. niger* (GaaX). AnGaaX was shown to repress activity of GaaR in the absence of GalA (39). Although we did not genetically explore the conservation of function of this repressor, the function of GaaX is likely conserved between species. The homolog of AnGaaX in *N. crassa* is NCU04298. NCU04298 and AnGaaX have 50% amino acid identity and 64% of amino acids are positive matches. Like GaaX, NCU04298 is also clustered in the genome with *pdr-2.* The binding motif of GaaR is present in the promoter region of GaaX, indicating that GaaX is positively regulated by GaaR. The authors of the study hypothesize that this is a way ensure pectin catabolism is tightly controlled as there is always enough GaaX to repress the amount of GaaR present. In *N. crassa,* we also see evidence of positive regulation of NCU04298 by Pdr-2. Our differential expression data between ΔPdr-2 and WT show that expression of NCU04298 is ~16 fold reduced in the ΔPdr-2 strain. Due to the conservation of these attributes between NCU04298 and GaaX, we hypothesize that GaaX is conserved in function between *A. niger* and *N. crassa.*

We were unable to find any additional pectin related TFs. PDR-1 and PDR-2 are likely the major TFs involved in pectin metabolism, and their regulons cover all of the pectinases and GalA and rhamnose metabolism genes required for catabolism of the pectin backbone.

4.3. Arabinose and arabinan metabolism is governed by previously characterized and novel TFs

4.3.1. ***Introduction***

Arabinose is an abundant sugar in the plant cell wall, and is a viable source of energy for filamentous fungi. It is a five carbon sugar is found primarily in hemicellulose polymers, most abundantly as residues present on the side chains of xylan and xyloglucan (164). It is also present in the side chains of pectin polymers rhamnogalacturonan I and homogalacturonan (165). Depending on the plant species, arabinose composes up to ~50% of the sugar content in the hemicellulose portion of the plant cell wall, and up to ~15% of total plant cell wall (166). This makes arabinose a non-trivial sugar for organisms utilizing the plant cell wall for metabolic energy.

Arabinose metabolism has been studied in several filamentous ascomycete fungi and the fundamental mechanism of its break down is well conserved. Arabinose catabolism has been investigated in *T. reesei* (167), *A. niger* (168) and *Pyricularia oryzae* (41) and others. In these fungi, arabinose is converted to L-arabinitol by L-arabinose reductase, then to L-xylulose by L-arbinitol dehydrogenase (LADH) and finally to xylitol by L-xylulose reductase. Once converted to xylitol, this intermediate joins paths with xylose catabolism pathway that eventually takes these sugar intermediates to the pentose phosphate pathway (169). Interestingly, *Trichoderma* and *Pyricularia* have a single reductase enzyme capable of converting both arabinose and xylose to their reduced sugar alcohol forms, whereas *A. niger* has two unique enzyme that catalyze each reaction independently. It is possible that the additional enzyme in *Aspergillus* allows for more specific regulation of arabinose metabolism. *N. crassa* appears to also have a single xylose/arabinose reductase like the closely related *Trichoderma* and *Pyricularia*.

The fungal enzymes involved with freeing arabinose from polysaccharides have also been well studied. The two major types of enzymes responsible for depolymerizing arabinan extracellularly are endoarabinases that cleave within the polysaccharide backbone and α-arabinofuranosidases that free arabinose from the ends of polymers (169). Both of these types of enzymes are present in the *N. crassa* genome. It has four α-arabinofuranosidases (NCU09170, NCU05965, NCU02343, NCU09775), one endoarabinase (NCU00852), one xylosidase/arabinosidase (NCU01900) and a rare exoarabinase from glycosyl hydrolase family GH93 (NCU09924). An additional L-arabinose transporter (NCU02188) has recently been described (96) (Table 4-3). The exoarabinase (NCU09924) has not been tested for activity in *N. crassa,* however, exoarabinase activity has been shown in homologous proteins (46% AA identity) from the closely related species *Fusarium graminearum,* and *Penicillium chrysogenum* (169, 170).

Gene ID	**Annotation**	**Extra or Intra-cellular**
NCU09170	α-N-arabinofuranosidase	extra
NCU05965	α -L-arabinofuranosidase	extra
NCU02343	α -L-arabinofuranosidase	extra
NCU09775	α-N-arabinofuranosidase	extra
NCU00852	endoarabinase	extra

NCU01900	xylosidase/arabinosidase	extra
NCU09924	Exoarabinase GH93	extra
NCU02188	L-arabinose transporter LAT-1	intra
NCU00643	L-arabinitol-dehydrogenase LADH	intra
NCU08384	xylose/arabinose reductase	intra
NCU09041	xylulose reductase	intra

Table 4-3 Arabinose and arabinan metabolism genes

Two transcription factors have been found in Ascomycota shown to specifically regulate arabinan/arabinose metabolism: AraR and Ara1. AraR is a positive regulator of arabinose pathway and is associated with extracellular arabinose production in *Aspergillus* species and other Eurotiales (40, 171). This TF is not present in Sordariomycetes, but an analogous TF, Ara-1 has been found recently in *P. oryzae* (41). Ara-1 has been shown to positively regulate the enzymes in the arabinose pathway as well as at least one arabinofuranosidase. Prior to this research it was not known whether the role of Ara-1 was conserved in *N. crassa*. It was previously shown that several arabinases namely the xylosidase/arabinosidase NCU01900, and the L-arabinofuranosidase NCU02343 are regulated by the TF XLR-1. These two genes are directly bound by XLR-1 and show far reduced expression in the Δ*xlr-1* strain (31). Additional arabinases are affected by XLR-1 but whether they are directly regulated by the TF is unclear.

We found that the homolog of *Ara-1* was present in *N. crassa* (NCU05414) as a TF induced by a number of different plant biomass, and suppressed under sucrose conditions (Figure 4-4).

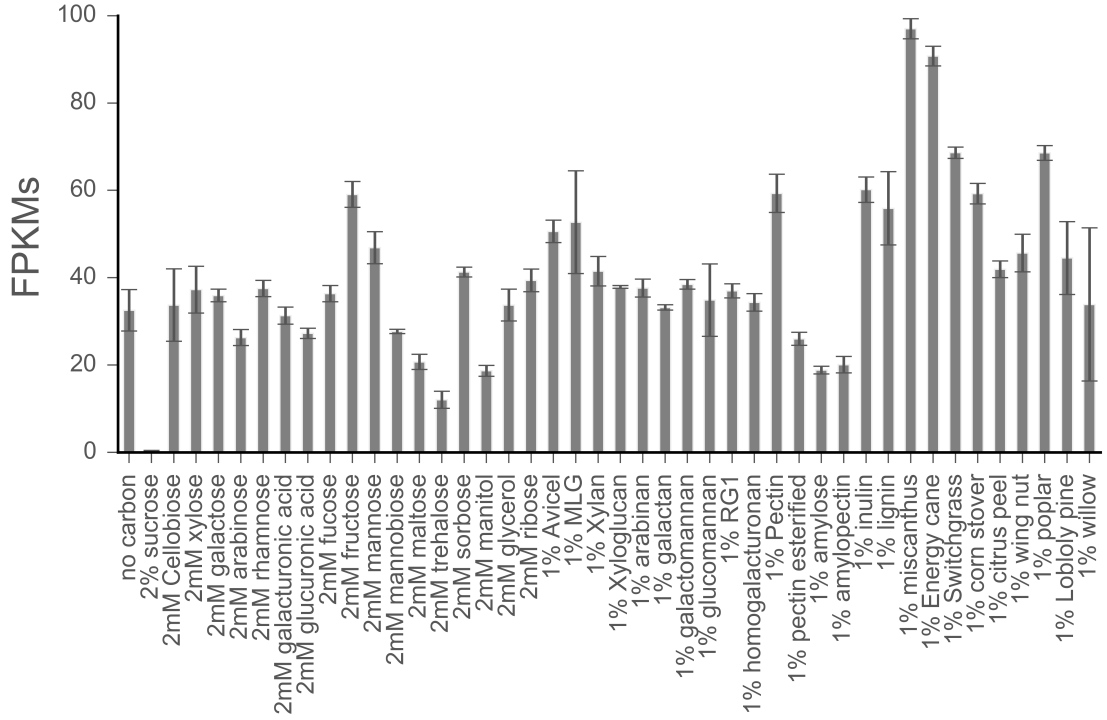


Figure 4-4 Expression of NCU05414 across carbon conditions. *Expression as measured by FPKMs with error bars representing standard deviation from triplicate.*

By conducting RNAseq on the ΔNCU05414 strain grown on Miscanthus and comparing to WT (FGSC 2489), we show that NCU05414 positively regulates the arabinose pathway, two arabinofuranosidases, an L-arabinose transporter, a β xylosidase and a number of additional genes. We further confirm the critical role of NCU05414 in arabinose metabolism by examining growth of WT, ΔNCU05414 and a strain overexpressing NCU05414 (OxNCU05414) in arabinose conditions. We additionally show with qRT-PCR that NCU05414 is required for LADH expression, which we believe to be the major cause for ΔNCU05414's reduced growth on arabinose. We name this transcription factor *ara-1* for its homology to *A. niger* and *P. oryzae Ara-1.*

4.3.2. ***Results***

Arabinase gene expression levels may distinguish metabolic importance.

N. crassa has four arabinosidases, one endoarabinase, one xylosidase/arabinase, and a GH93 exoarabinase. Of these enzymes, arabinosidase (NCU02343), xylosidase/arabinase (NCU01900), arabinosidase (NCU09170) and arabinosidase (NCU09775) display the highest levels of transcripts. Arabinase (NCU09924), and arabinosidase (NCU05965) show moderate expression and endo-

arabinase (NCU00852) shows extremely low expression levels (Figure 4-5). This indicates that NCU00852 is likely not a crucial enzyme for freeing arabinose from the plant cell wall, while the other enzymes likely have a more important role in this process. It is possible that we did not test any conditions or time points to fully induce NCU00852.

The conditions each enzyme is induced on also clue us into the type of polysaccharides that the enzyme likely degrades. Arabinases NCU09170, NCU09775 and NCU05965 are only induced highly on pectin, homogalacturonan, and citrus peel conditions (Figure 4-5A,C,G). These data indicate that these three enzymes are likely involved in arabinose release from pectin. Conversely, xylosidase/arabinase NCU01900 is induced more by xylan and the grass plant biomasses, which contain mostly cellulose and hemicellulose (Figure 4-5E). Arabinosidase (NCU02343) and arabinase (NCU09924) are induced by both pectin and hemicellulose, indicating that these enzymes likely work on both types of polysaccharides (Figure 4-5B, D).

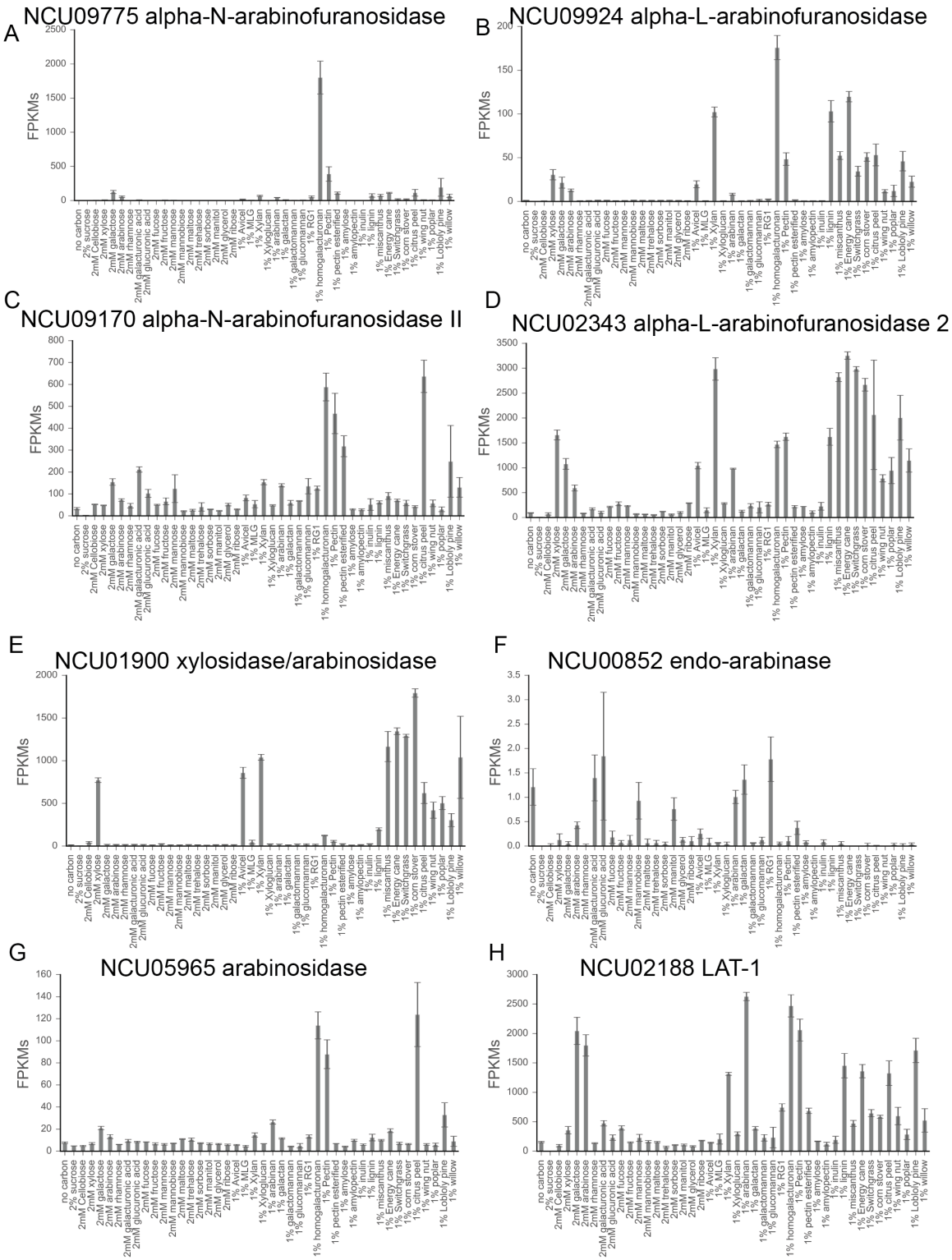


Figure 4-5 Expression of arabinases in WT across carbon conditions. ***A-G*** *display the normalized counts in FPKMs across all carbon conditions.* ***H*** *displays normalized counts of the known arabinose specific sugar transporter LAT-1. Error bars represent standard deviation from biological triplicates of each condition.*

Positive regulation of Arabinases by TFs

Several previously described and several newly described *N. crassa* TFs affect expression of arabinases. Key transcription factors in *N. crassa* that regulate extracellular carbon metabolism that have already been described include cellulase regulators *clr-1* and *clr2*, hemicellulase regulator *xlr-1*, and pectinase regulator *pdr-1* (27, 30, 35). We examined the expression of arabinases in TF deletion and WT strains with RNAseq. We were able to replicate previously published data showing that XLR-1 and PDR-1 affect expression of several arabinases (Figure 4-6A and B). Deletion of *clr-1* and *clr-2* has minimal effect on the expression of most arabinases, except for arabinosidase NCU09775. Both *Δclr-1* and *Δclr-2* strains show ~10 fold expression difference for NCU09775 under Avicel conditions, however the normalized expression for this gene under Avicel are relatively low compared to higher NCU09775 inducing conditions such as xylan or grassy plant biomasses (Figure 4-6C and D). These cellulase regulators could possibly affect expression of NCU09775 under high inducing conditions as well. XLR-1 contrastingly exerts positive regulation on much larger number of arabinases. Deletion of *xlr-1* most greatly affects expression of arabinosidases NCU02343, NCU01900, and NCU09924. It also affects expression of NCU09775 but to a lesser extent than the cellulase regulators. PDR-1 seems to have miniscule affects on the arabinases, but does display some control over LADH crucial for intracellular arabinose metabolism.

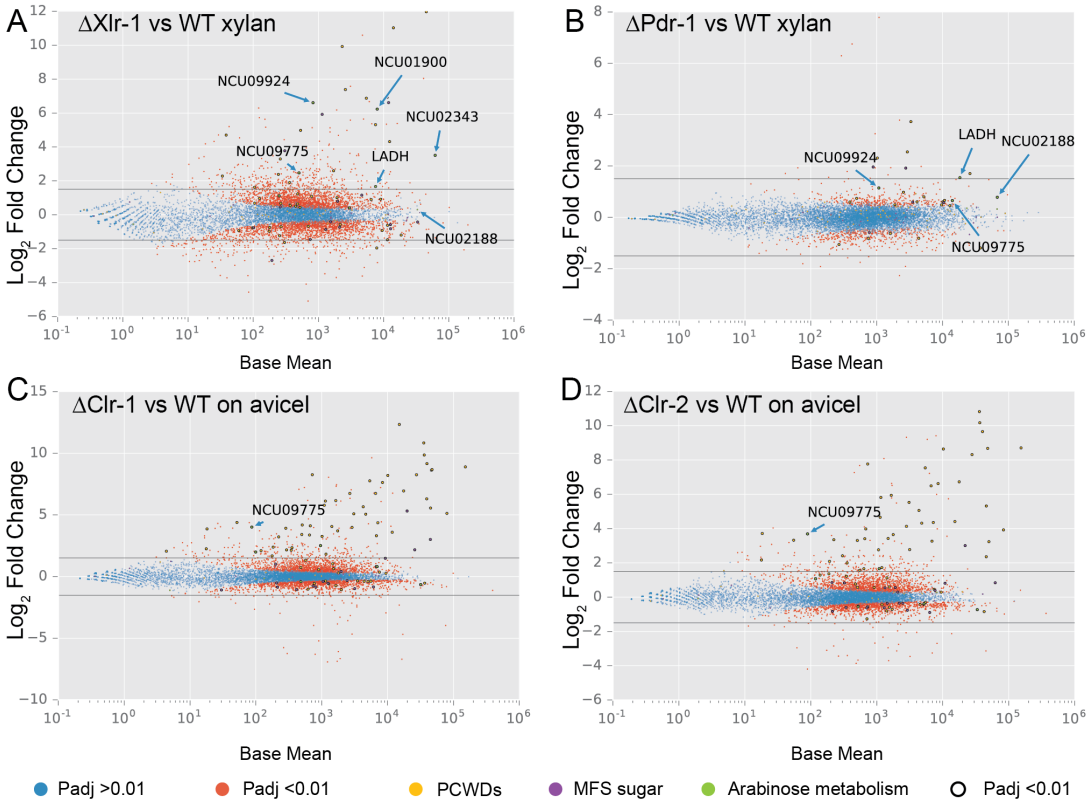


Figure 4-6 Transcriptional regulation of arabinases (indicated) by previously described carbon metabolic TFs. *Differential expression as data calculated by*

DEseq plotted by log_2 fold change vs base mean for Δxlr-1 (A), Δpdr-1 (B), Δclr-1 (C) and Δclr-2 (D) vs WT where ΔTF are the reference conditions. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj= 0.01.

We also examined transcription factors that have not yet been explored in *N. crassa* and their role in arabinose metabolism. Some of these have homologs already described in closely related lineages, such as *ara-1* homolog to *Ara-1* in *P. oryzae* and *pdr-2* homolog to *GaaR* in *Botrytis cinerea* (37, 41). ARA-1 displays a clear regulatory role over arabinosidase (NCU09775), LADH and LAT-1 (Figure 4-7C). It also shows some control over arabinosidase (NCU09924) but perhaps less than XLR-1 and NCU01074 if comparing by fold change of expression. ARA-1 is unique in its regulatory affects on LADH and LAT-1 both of which are crucial for intracellular arabinose metabolism. PDR-2 shows a moderate regulatory role for the regulation of expression of arabinases in *N. crassa.* Deletion of *pdr-2*, reduces expression levels of arabinosidases NCU09775, NCU05965 and NCU09170 dramatically (Figure 4-7A). These arabinosidases are the same that are highly up regulated on pectin rich conditions, further supporting the conserved role of NCU04295 across plant cell wall degrading species in the Ascomycota.

NCU01074 is another previously undescribed TF that appears to positively regulate arabinase gene expression. This TF is a basic leucine zipper. Deletion of NCU01074 reduced expression of arabinosidases NCU09775 and NCU09924, and to a lesser extent, NCU09170. It may also play a minor role in positive regulation of the two LADH and LAT-1 as well as the arabinosidase NCU02343 (Figure 4-7C). In addition to arabinases, NCU01074 also affects expression of number of xylanases including β- xylosidases (NCU09923, NCU00709), endoxylanases (NCU05924, NCU07225), and acetylxylan esterases (NCU09664, NCU04870). These xylanases are down regulated between 4 and 16 fold in the ΔNCU01074. For its role in xylanase and arabinase expression we have named this TF *hem-1* for hemicellulase regulator -1.

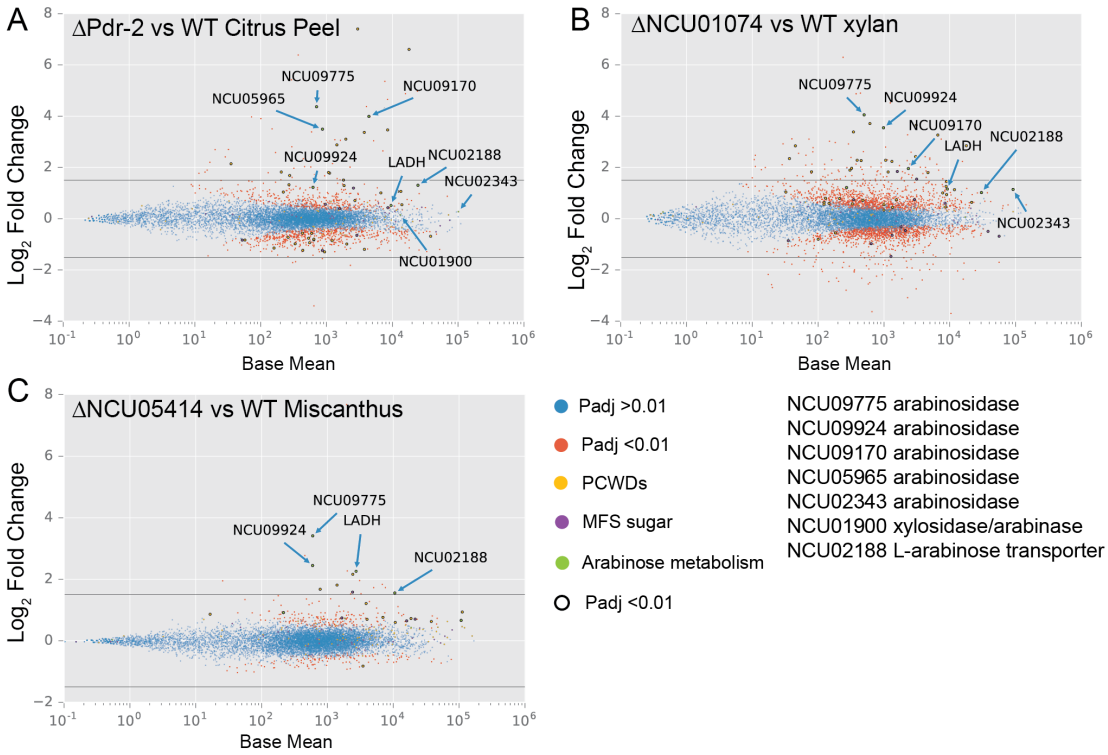


Figure 4-7 Transcriptional regulation of arabinases by previously undescribed TFs in *N. crassa*. *Differential expression as data calculated by DEseq plotted by log_2 fold change vs base mean for Δpdr-2 (A), Δhem-1 (NCU01074) (B), Δara-1 (NCU05414) (C) vs WT where ΔTF are reference conditions. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj= 0.01.*

TF NCU05414 homolog to Ara-1 plays critical role in arabinose metabolism.

We further explored the role of *ara-1* (NCU05414) due to its recent discovery in *P. oryzae* and as well as the results showing deletion of *ara-1* greatly reduces expression of LADH and LAT-1. We constructed a new strain (OxNCU05414) overexpressing *ara-1* under the glyceraldehyde phosphate dehydrogenase (GPD) promoter at the *csr-1* locus. When we compared growth of WT, ΔNCU05414 and OxNCU05414 on 2% arabinose, we observed that deletion of NCU05414 greatly reduced growth and overexpression of NCU05414 rescued growth (Figure 4-8A). Furthermore, NCU05414 was able to grow faster than WT on arabinose indicating that, expression of genes controlled by NCU05414 was restricting WT from growing at faster rates. We further show that LADH was overexpressed in the OxNCU05414 strain under arabinose by qRT-PCR (Figure 4-8B). These data suggest that LADH expression levels can increase arabinose utilization. It is however possible that LAT-1 expression is the true bottleneck and further experiments must be conducted to determine why OxNCU05414 is more capable of utilizing arabinose.

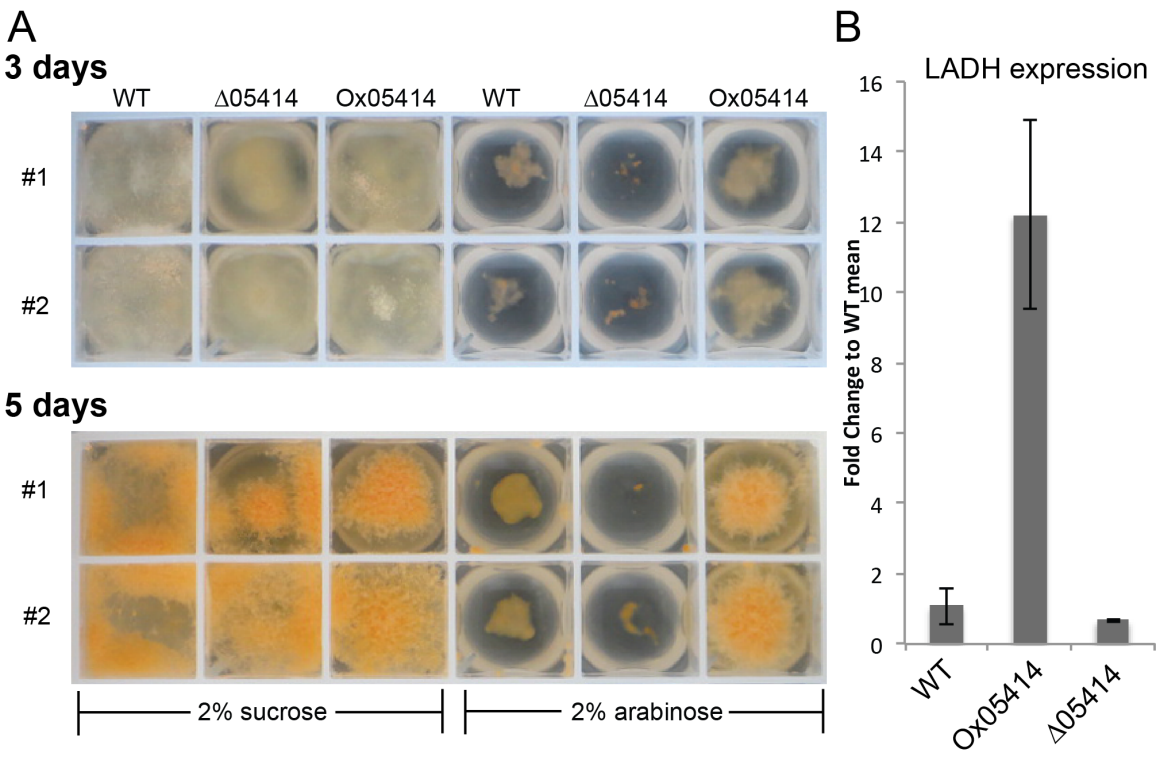


Figure 4-8. Growth and LADH expression of *ara-1* deletion (*ΔNCU05414*) and *ara-1* overexpression (OxNCU05414) strains. *A) Growth at 3 day and 5 day 3ml liquid cultures with 2% arabinose as sole carbon source. Duplicates shown in 2nd row of each panel. B) Expression of LADH as measured by qRT-PCR after 4 hour shift to 2% arabinose media post 16 hour growth in Vogel's MM (2% sucrose). Experiment conducted in biological triplicate, and presented as fold change relative to WT triplicate mean expression. Error bars represent standard deviation of ΔΔCt values.*

4.3.3. *Conclusions*

We find that arabinose metabolism is regulated by a number of TFs including some previously described and some that had yet been explored in *N. crassa*. Extracellular degradation of hemicellulose and pectin can free arabinose that can be transported and metabolized by the cell. Enzymes involved in the release of arabinose are limited in *Neurospora* and regulated in a joint manner by many transcription factors including *xlr-1*, *clr-1*, *clr-2*, *pdr-2*, *hem-1* (NCU01074) and *ara-1* (NCU05414). We show that single deletions of several different TFs can affect expression of the same arabinase gene under different conditions. Expression of genes involved in intracellular arabinose metabolism is also affected by several different transcription factors, but the TF *ara-1* (NCU05414) is crucial. This TF is required for full expression of LADH, which is necessary for the conversion of arabinose to xylulose, and entry into the pentose phosphate pathway. The function of ARA-1 is conserved between *N. crassa*, *P. oryzae* and *T. reesei*, as the growth phenotype observed in ΔNCU05414 strain is the same in all three species (41).

We also discover a novel arabinase regulator, *hem-1*, that also regulates a subset of xylanases. *Hem-1* expression is highest under 2mM xylose, indicating that activity of HEM-1 may be linked to the presence of that sugar. Interestingly, HEM-1 is not a zn2cys6 TF. All of the positive regulators of PCWDEs that have been discovered are all Zn2Cys6 TFs, which is what makes this discovery surprising. *Hem-1* is also not well conserved outside of Sordariomycetes. It is present in *Trichoderma, Sordaria* and *Fusarium,* but is not present in most *Aspergillus* species except for *A. fischeri.*

4.4. Minor carbon response regulators

4.4.1. *Introduction*

From our candidate transcription factors we tested (see section 3.4), we found several additional TFs that displayed modulated expression of PCWDEs unrelated to pectin or arabinan utilization. These data indicate that although *N. crassa* has primary TFs, it simultaneously utilizes a number of minor TFs for further modulation of expression of these enzymes. Many of these TFs also affect expression of unique sets of genes where it is difficult to discern whether there is a common function or pathway that theses genes are a part of. These gene sets are often not enriched for GO categories or functional categories. In this section, we explore the function of some of these minor TFs.

The data presented in this section needs much further confirmation. Most of the conclusions drawn are based only on RNAseq, which can sometimes result in erroneous conclusions. The data sets presented here are either very noisy such that the likelihood of experimental/technical failures are higher than other datasets or differential expression levels are too low to draw strong conclusions. One issue we faced, especially with solid substrates like full plant biomasses and Avicel, was inconsistent growth of mycelia on the substrate. Variations on mycelial mat fluctuated day to day. Sometimes, mycelia easily grew and surrounded the substrate but other times would stay suspended and make limited contact with substrate. This data is included in the dissertation mainly because it summarizes the work we conducted that may be useful to future researchers that come across these same TFs.

4.4.2. *Minor regulators of PCWDE expression*

NCU03421 appears to be minor regulator of genes encoding hypothetical proteins and PCWDEs

NCU03421 is annotated as "zinc finger transcription factor 23." No research has been published on this TF in any fungi. The TF does not appear to be well conserved across the fungal kingdom. We found that it was up regulated on many different plant biomass. These data indicated that a specific compound present in the plant biomass was inducing the up regulation of this TF. We also concluded that the complex polymers did not contain the signal, except for three of the polysaccharides tested. Low-level up regulation is noted on xyloglucan, pectin and

glucomannan. The differential expression data shows a number of LPMOs and β-1,4 glucanases up regulated 2-6 fold in the WT Strain. We checked to see whether *clr-2* expression was affected as these PCWDEs are all part of the CLR-2 regulon. Indeed, expression of *clr-2* is ~2x decreased (adj P = $4x10^{-9}$) in the ΔNCU03421. We do not have any further evidence for NCU03421 regulation of *clr-2*.

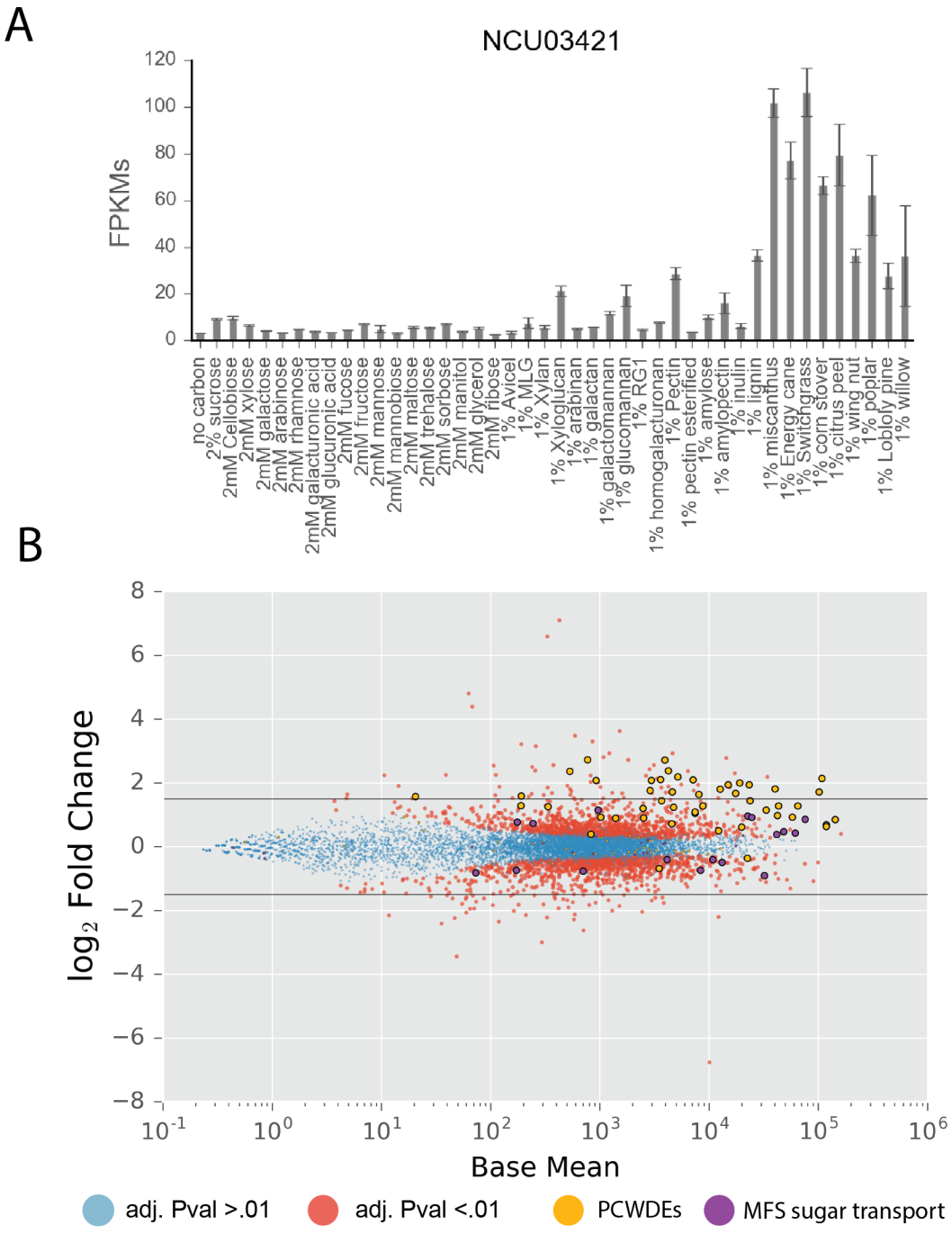


Figure 4-9 NCU03421 expression across carbon conditions and differential expression analysis of ΔNCU03421 vs WT. *A) expression as measured by FPKMs of*

NCU03421 across carbon conditions. Note increased expression under Miscanthus, energy cane and fucose. Error bars represent standard deviation of biological triplicates. B) Differential expression of NCU03421 vs WT with ΔNCU03421on Miscanthus as references condition. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj= 0.01.

NCU01312 *rca-1* may be negative regulator of PCWDEs

NCU01312 *rca-1* is a previously described MYB like helix-turn-helix TF that appears to repress expression of PCWDEs. The TF has been investigated previously for its role in conidiation (172). The acronym RCA stands for "regulator of conidiation in *Aspergillus.*" A previous study found that the homolog of *rca-1* in *A. niger (flb-1)* is required for WT conidiation. *Rca-1* from *N. crassa* complements the deletion of *flb-1*, but deletion of the gene in *N. crassa* did not affect conidiation. However, an interesting phenotype of the *Δrca-1* mutant was observed. When grown on a petri dish, WT *N. crassa* normally forms a clockwise spiraling colony. This was not observed in *Δrca-1* (172). We found this TF due to its slight up regulation on Miscanthus and energy cane. We conducted RNAseq with *Δrca-1* on 1% Miscanthus. Our results showed that many PCWDEs were down regulated in the WT strain, the majority being a part of the CLR-2 regulon. Indeed *clr-2* is ~3.5 fold increased in the *Δrca-1* mutant. Results indicate that RCA-1 is likely a repressor of *clr-2* and possibly some of the cellulase genes, in addition to its previously described role in spiral-shaped hyphal growth.

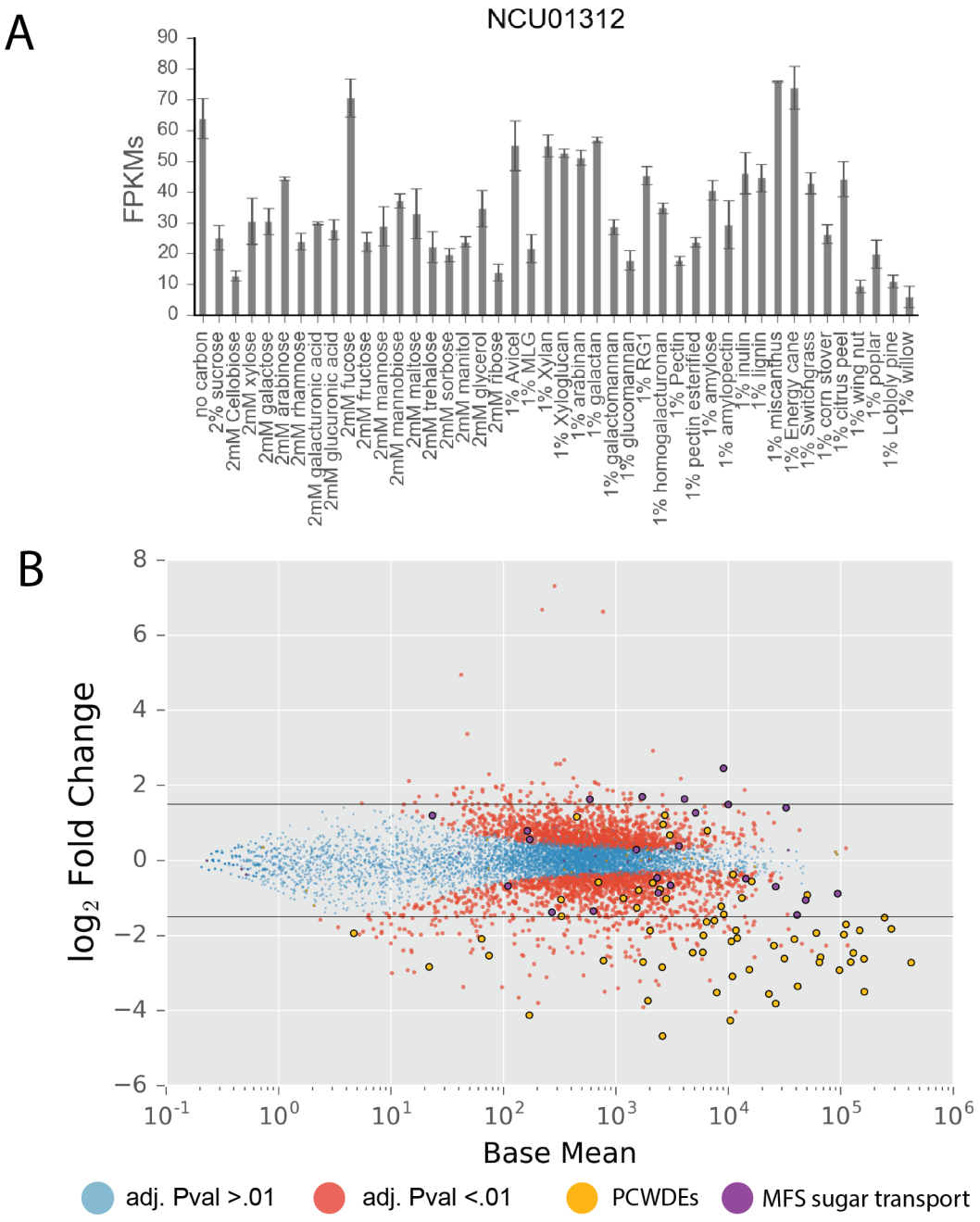


Figure 4-10 NCU01312 (*rca-1*) expression across carbon conditions and differential expression analysis of *Δrca-1 vs* WT. *A) expression as measured by FPKMS of rca-1 across carbon conditions. Error bars represent standard deviation of biological triplicates. B) Differential expression of Δrca-1 vs WT on 1% Miscanthus with Δrca-1 as reference condition. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj= 0.01.*

NCU00445 *man-1*

In our search for novel transcription factors, we hypothesized that *N. crassa's* single extracellular β 1-4 mannanase (NCU08412) and two β-mannosidases (NCU00890, NCU00985) that are up regulated under mannan conditions might be under control of a mannose specific TF. We therefore looked for TFs that were up regulated on 2mM mannose or mannan derivatives. We found a TF NCU00445 that was slightly up regulated on 2mM mannose (Figure 4-11A). We conducted RNAseq on the ΔNCU00445 on 2mM mannose and compared gene expression to WT on mannose using DEseq. We found that a large number of genes were differentially expressed between the two strains. We also saw a number of PCWDEs and MFS sugar transporters that are up regulated in WT as compared to ΔNCU00445. β-1,4 mannanase (NCU08412) was the top differentially expressed PCWDE gene with ~16 fold change. These data indicated to us that it was under regulation by NCU00445. We found that a number of other PCWDEs are also up regulated in WT including feruloyl esterase B (NCU09491), acetyl xylan esterase (NCU04870) and many other PCWDEs but at low fold change (log2foldchange ~ 2). However, the base mean for these genes was also pretty low indicating that mannose is a very low inducing condition for these genes to begin with. Regulation of these genes may be more relevant in other conditions where induction of these PCWDEs is higher. We named NCU00445 *man-1* for "mannanase regulator 1." The data for *man-1* vs WT is however very noisy as demonstrated by the spread of the scatter plot (Figure 4-11 B). Thus, it is possible that genes expressed with low differential expression may be due to noise, or technical aspects of the experiment that were not identical between *Δman-1* and WT. The direct binding sites of MAN-1 are further explored in section 5.4, and reveal a more specific role of the MAN-1 TF.

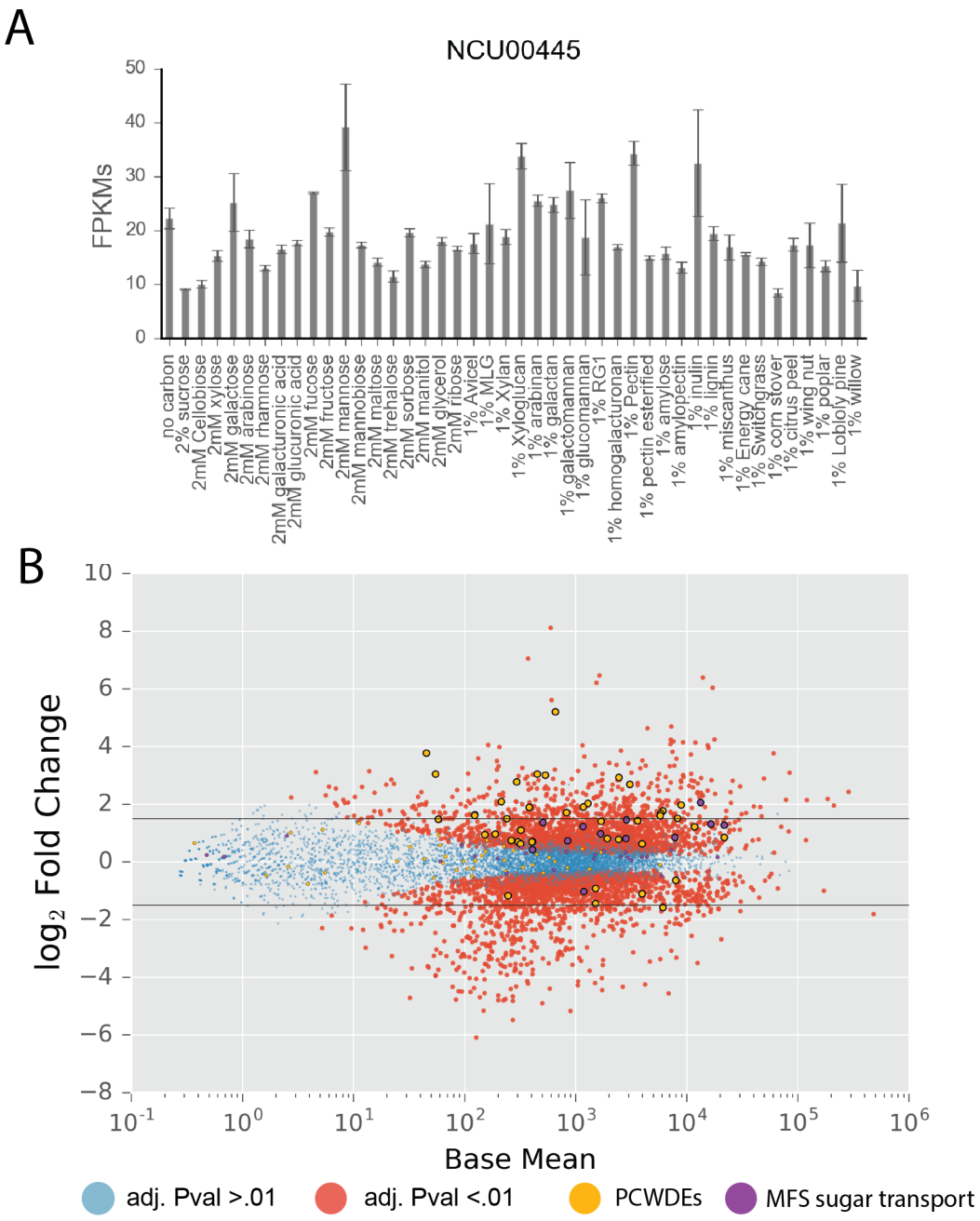


Figure 4-11 *Man-1* (NCU00445) expression across carbon sources and differential expression analysis of Δ*man-1 vs* WT. *A) expression as measured by FPKMS of NCU00445 across carbon conditions. Note increased expression under 2mM mannose. Error bars represent standard deviation from biological triplicates. B) Differential expression of ΔNCU00445 vs WT with ΔNCU00445 as the reference strain. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj= 0.01.*

4.4.3. ***Regulators that show differential expression but of unknown function.***

We found three additional transcription factors that showed differential expression between ΔTF and WT. The differentially expressed genes do not include many PCWDEs. The gene sets that are differentially expressed are not enriched for GO terms, and are often include genes encoding hypothetical proteins. For example, of the top 50 genes down regulated in the TF deletion ΔNCU04058 versus WT under 2mM ribose, ~70% of genes encode hypothetical proteins. The annotated genes fit no category or overarching function as of yet. We present these TFs that appear to affect gene expression but we are unable to ascertain a possible function.

The four TFs presented here include NCU00282, NCU04058, NCU03471 and NCU03699. NCU00282 is a Zn2Cys6 TF with homologs scattered within Sordariomycetes. NCU04058 is a bZIP TF that is well conserved among filamentous ascomycete species but not *Saccharomyces*. NCU03417 and NCU03699 are both well-conserved Zn2Cys6 TFs. None of these TFs have any annotated function. All of them show unique transcriptional profiles, and are up regulated on one or several carbon conditions (Figure 4-12).

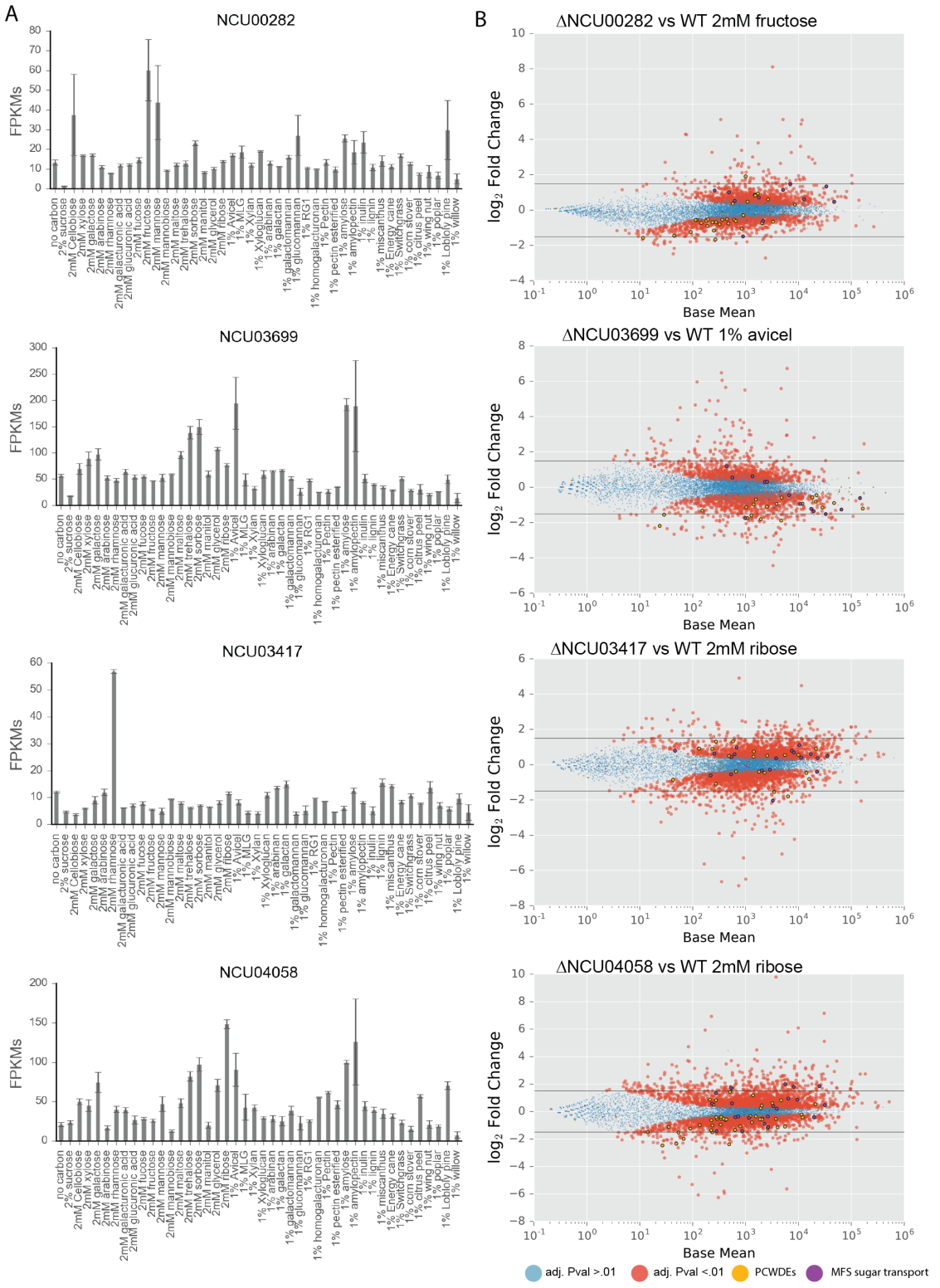


Figure 4-12 Expression and differential expression analysis for TFs with non-PCWDE modulating regulons. *A) Expression as measured by FPKMS of NCU00282, NCU04058, NCU03471 and NCU03699 across carbon conditions. Error bars denote standard deviation from biological triplicate. B) Differential expression of*

ΔNCU00282, ΔNCU04058, ΔNCU03471 and ΔNCU03699 with ΔTF as reference strains. Horizontal lines added to each graph represent + and – 1.5 log_2 fold change but have no further significance and are present only for better comparison across plots. Data points that have P adjusted values below than 0.01 have additional outlined circles to distinguish from those above Padj= 0.01.

4.5. Summary of known direct PCWDE regulators and what they reveal about PCWDE production

In this chapter we showed that RNA profiling can be used to find additional TFs responsible for regulating expression of PCWDEs. However, it is not the most efficient way of searching. We found and tested 37 candidate transcription factors for an effect on PCWDE expression. Four of these modulate PCWDE expression to varying extents: NCU04295 (*pdr-2*), NCU05414 (*ara-1),* NCU01074 (*hem-1*) and NCU00445 (*man-1).* An additional two TFs NCU01312 and NCU03421 show some sign of PCWDE modulation, but results need to be verified. A number of additional TFs (we present 4 here) appear to modulate expression of genes, but the processes they regulate are unclear. They do not affect expression of PCWDEs genes significantly.

In retrospect, the important PCWDE modulating TFs we found were all within our group of most promising candidates. They were also some of the first we tested. Thus, it may not be worthwhile to continue searching for TFs, when TFs with interesting profiles are exhausted. Another key aspect of our search was filtering TFs for only ones with reduced expression on sucrose. Logically, TFs with reduced expression on sucrose are likely under CCR, and thus more likely to be PCWDE regulators or some form of carbon metabolism regulator. The successful candidates all turned out to have reduced expression under sucrose. With this knowledge in hand, it might be easier to conduct a similar search for TFs in another organism with far more efficiency. With the findings presented here along with recent discoveries of TFs that regulate PCWDE genes in other ascomycete fungi, novel TFs that greatly impact PCWDE expression are likely all discovered. With the present information, we can explain the production of most of the enzymes required to break down the primary components of the plant cell wall: cellulose, hemicellulose and pectin.

Lignin is the final plant polymer that we lack an understanding of the transcriptional regulation responsible for its deconstruction. We found a TF greatly induced on lignin, NCU02142, but deletion of the TF yielded very minor expression changes on lignin relative to wild type cells at the 4 hour time point (data not shown). *N. crassa* cannot utilize lignin, but we reasoned that lignin may be toxic to the cell and that the TF may regulate some form of cellular detoxification. Transcriptome profiling in a lignin-degrading organism may reveal a conserved regulator for lignin depolymerization. The strategy we used could applicable for the discovery of such a regulator.

Regarding *pdr-2* and *ara-1,* much of the work was halted when the characterization of these TFs were published by other labs. We found that our data matched results found in other organisms, and that these TFs are well conserved in function. Data regarding *hem-1* (NCU01074) remains unpublished in other

organisms, perhaps due its limited conservation among closely related Sordariomycetes.

Our analysis of multiple TFs allowed us to gain a much better understanding of the complexity of PCWDE gene expression. The regulons of the major PCWDE regulators, CLR-1, CLR-2, XLR-1, PDR-1, PDR-2, ARA-1, HEM-1, MAN-1, COL-26 all show extensive overlap. To visualize this overlap, we constructed a Circos (173) plot, showing all of the PCWDE transcription factors, and their positive regulation of PCWDEs (Figure 4-13). Each line connecting a transcription factor to a PCWDE or other gene represents DEseq data between WT and the ΔTF mutant strain. The thickness of each line represents the $\log_2$ fold change of differential expression. This depiction provides a proxy for the amount of positive control that a TF has over a gene. As we can see, many TFs regulate expression of the same PCWDEs genes in different conditions. Overlapping regulation was also observed by the analysis of 2mM sugars as discussed in section 3.3.3 and can be explained by our TF data.

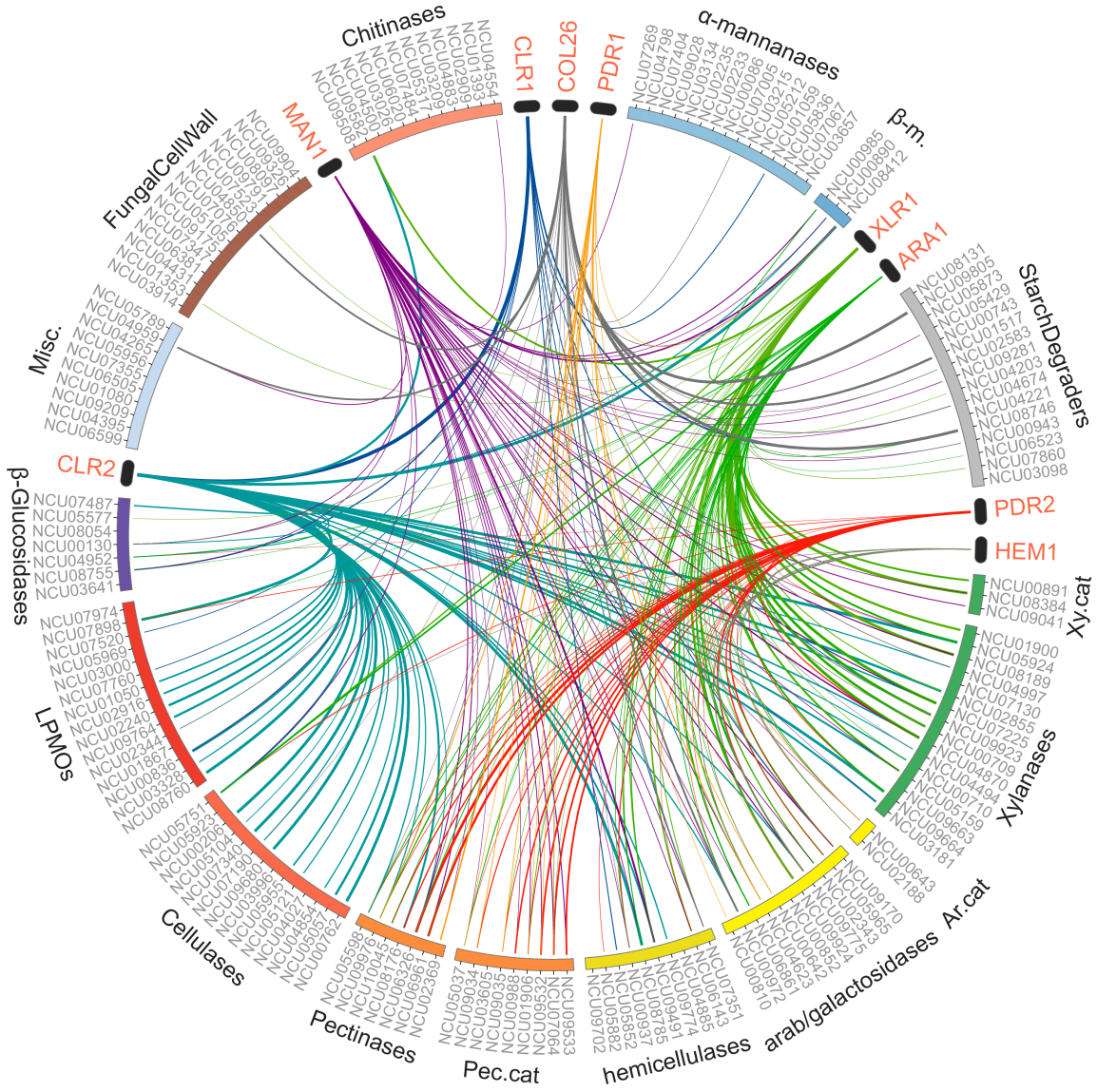


Figure 4-13 Overlapping regulons of major PCWDE regulators. *We used Circos v0.69 to build this figure* (173)*. We have divided catabolic CAZymes into functional groups displayed on the outer edge of the plot. Each CAZyme is represented by its gene ID. "Pec. cat" is shorthand for pectin catabolism, "Ar" arabinose, "Xy" xylose and "m" for mannanase. Transcription factors are displayed by black ovals along the edge of the plot annotated with red font. Each line represents data from DEseq between WT and a ΔTF strain on the conditions described in the following: Δclr-1, Δclr-2 on 1% Avicel, man-1 on 2mM mannose, Δcol-26 on 2mM maltose, Δpdr-1 on 1% pectin, Δpdr-2 on 1% citrus peel, Δara-1 on 1% Miscanthus, and Δxlr-1 and Δhem-1 on 1% xylan.*

Surprisingly, none of the TFs found seemed to modulate expression of other TFs. This excludes the CLR-1-CLR-2 interaction described previously (27, 31). This information along with information regarding their activation indicates that each of these TFs have a generally independent signal for induction and activation.

CLR-1 and 2 regulation of PCWDE expression is also unique among the major TFs in another manner. Cellulase/ LPMO expression is almost solely the responsibility of these TFs. With other sets of PCWDEs genes, we observe multiple TFs activated by different sugars that induce expression. Contrastingly, cellulases/LPMOS are really only under the positive regulation of CLR-2. Simultaneously, CLR-2 regulates a large number of non-cellulose PCWDEs in addition to cellulases. We are not sure why regulation of cellulases happens in this manner. It is possible that are instances where cellulose is absent from the substrate that the cell is consuming thus it would be disadvantageous to produce those enzymes. Or perhaps, because cellulose is so costly to depolymerize due to its recalcitrant structure, the cell would rather focus on other accessible polymers initially. The up regulation of these PCWDEs genes that depolymerize accessible polysaccharides without production of cellulases could be a frequent mode of action for *N. crassa.* Investigation of this topic in other organisms will be helpful in determining the cause of this phenomenon.

Meanwhile, expression of hemicellulases and pectinases is regulated by a large number of TFs. Almost all the TFs we describe modulate expression of both hemicellulases and pectinases. Enzymes required for sugar catabolism are more specific to a particular TF. For instance, xylose metabolic genes xylose reductase and xylitol dehydrogenase are primarily regulated by XLR-1. GalA and rhamnose catabolic genes LRA1-4 and GaaA-D are mostly regulated by PDR-1 and PDR-2. However, overlapping regulation is still present for these genes (see pectin catabolism Figure 4-13). Pectin and hemicellulose are much more complex and are also diverse substrates compared to cellulose. Perhaps complexity and diversity of these polymers drives the necessity for more fine-tuning of expression of PCWDEs. Allowing multiple TFs to modulate expression of the same PCWDE can certainly allow for better control of expression. A simpler explanation is that pectin and hemicellulose are always found together in the substrates that *N. crassa* colonizes.

In summary, we find a system of regulation where independent TFs act in unison to modulate expression of many of the same PCWDEs, with cellulases/LPMOs as an exception. The majority of PCWDE regulating TFs do not modulate expression of one another, with the exception of CLR-1 and *clr-2*. Additionally, each PCWDE regulating TFs generally regulates catabolism of a single important sugar like xylose, GalA, rhamnose and arabinose.

4.6. Expanding mechanisms of phosphate metabolism in *N. crassa*

4.6.1. *Exploration of NUC-1 regulon shows that the TF is responsible for much more than previously thought.*

For more information regarding basic phosphate metabolism see section 3.6.1 regarding *N. crassa* response to starvation

RNAseq data recapitulates previous findings that NUC-1 is required for expression of several key phosphatases.

To explore the full regulon of NUC-1 we conducted RNAseq on *Δnuc-1* strain on no phosphate and compared expression profiles to that of the WT strain exposed to the same condition. Our data supports previous findings that several phosphatases are positively regulated by NUC-1 (Figure 4-14) (174–177). Four of these genes are highlighted in green and include alkaline phosphatase *pho-2* (NCU01376), acid phosphatase *pho-3* (NCU08643), inorganic phosphate transporters *pho-4* (NCU09564) and *pho-5* (NCU08325). One additional green highlighted gene is the glycerol diester phosphodieserase-1 (NCU10038) that has not been described within the NUC-1 regulon previously but is one of the most differentially expressed genes in our DEseq data. This enzyme cleaves glycerol phosphate from cell membrane phospholipids, a function that indicates a high probability that it is involved in phosphate scavenging.

Differential expression analysis identifies loci of characterized but unmapped nucleases and nucleotidase

Several nucleases have been previously described to be under the regulation of NUC-1 (139, 178). *Nuc-1* mutants were unable to express these nucleases. However, these enzymes were never mapped to gene loci in the modern genome sequence of *N. crassa.* To try and determine which nucleases these might be, we looked for nucleases that were down regulated in the *Δnuc-1* strain. In Figure 4-14 all nuclease genes are highlighted, but only three nucleases fit this description. These are ribonuclease T1 (NCU01045), nuclease PA3 *nuc-8* (NCU08648), and endonuclease/exonuclease/phosphatase *nuc-15* (NCU09788). All three nucleases have signal peptides indicating they are likely extracellular. Nuclease PA3 and ribonuclease T1 likely match with the ribonucleases first described in Hasunuma et al., 1976 and Lindberg and Drucker, 1984 (139, 178). We also found a 5'-nucleotidase (NCU09659) that is down regulated in *Δnuc-1* that we suspect is the same nucleotidase characterized and shown to be regulated by NUC-1, but never mapped to genome (138).

NUC-1 represses expression of ribosomal proteins.

We additionally observed a large set of genes that are up regulated when *nuc-1* is deleted. Interestingly, we noticed that the large cluster of ribosomal proteins that were down regulated in WT under phosphate starvation (Figure 3-19), are slightly up regulated in the ΔNuc-1 strain compared to WT (Figure 4-14). Although ribosomal proteins are not dramatically up regulated, the majority of them have ~2X more expression than in WT strain. Although many of the adjusted p-values are close to .01, the pattern that almost all the ribosomal proteins are up regulated illustrates to us that the observed phenomenon is authentic. These data indicate that either NUC-1 or likely a mechanism downstream of NUC-1 represses expression of ribosomal protein synthesis. This conclusion is further supported when we look at FunCat enrichment for genes with greater than 2.25 fold change between WT and *Δnuc-1*. The top 10 functional categories include ribosome biogenesis (#3), protein synthesis (#4) and ribosomal proteins (#11). The rest of

these functional categories are involved in transcription and RNA processing. These data indicate NUC-1 is involved in the switching off protein production during phosphate starvation through dampening of gene expression involved in transcription and translation. However, there must be other factors involved in ribosomal protein repression because the level of down-regulation seen in our WT phosphate vs no phosphate data is not reflected in the levels of up-regulation in *Δnuc-1* background (see Figure 3-19 and compare to Figure 4-14).

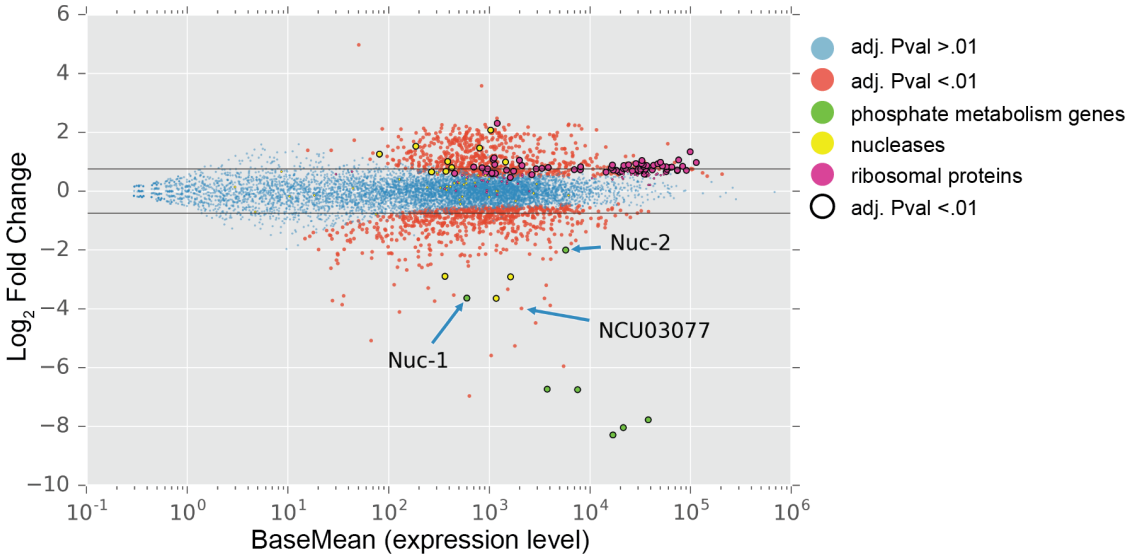


Figure 4-14 Differential Expression of WT vs *Δnuc-1* in no phosphate media. *Average expression of each gene is plotted along the x axis while log_2 fold change is plotted on Y with WT as the reference condition. Significantly differentially expressed genes with P adjusted values below 0.01 are represented in red, whereas statistically insignificant genes are represented in blue. Additional colors denote groups of genes of interest. Outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.01). Horizontal lines added to the scatter plot represent + and – .75 log_2 fold change.*

Nuc-1 crosstalk with carbon metabolism

We used FunCat to investigate additional genes that were down regulated in the *Δnuc-1* mutant along with the phosphatases and nucleases described above. These genes are therefore under positive regulation by NUC-1 during phosphate starvation. We were surprised to find that functional categories for hemicellulose and pectin metabolism were the top categories enriched (data not shown). Indeed, we find that *Δnuc*-1 strain has reduced expression of: *xlr-1* (NCU06971), *hem-1* (NCU01074), *Lat-1* arabinose transporter (NCU02188), high affinity glucose transporter and 3 of the 4 the key GalA catabolic genes GaaA-C (NCU09532, NCU09533, NCU07064). However fold changes only differ by 2.25X to 6X, which are not drastic changes in expression levels. Additionally, under sucrose conditions, these genes are lowly expressed to begin with, so final expression levels under *Δnuc-*

1 are not necessarily high in comparison to inducing conditions. These data do indicate that NUC-1 is either directly or indirectly affecting expression of a portion of the carbon catabolic genes during phosphate starvation. Although we do not have data for phosphate starvation under inducing conditions, it is very possible that combining limited phosphate under a carbon condition a full plant biomass could induce PCWDEs to even higher levels.

4.6.2. ***Identification of NCU03077 putative TF involved response to phosphate limitation***

To find additional TFs that are involved in the response to phosphate starvation we examined two groups of TFs: those differentially expressed when *nuc-1* is deleted and TFs differentially expressed under phosphate starvation. The first group of TFs should be downstream of NUC-1, where expression is completely or partially dependent on an active NUC-1 protein. The second group may additionally contain TFs independent from NUC-1 regulation. When we looked at both of these groups we saw that only 1 TF (NCU03077) had generally high expression under phosphate starvation specifically and also down regulated in the *Δnuc-1* strain, indicating that it acted downstream of NUC-1 (Figure 4-15A). Although we found other transcription factors from the differential expression analysis (NCU05767, NCU06965, NCU00282, NCU16599), these TFs show up regulation under a number of other conditions. Only NCU03077 was specifically up regulated under phosphate starvation (Figure 4-15).

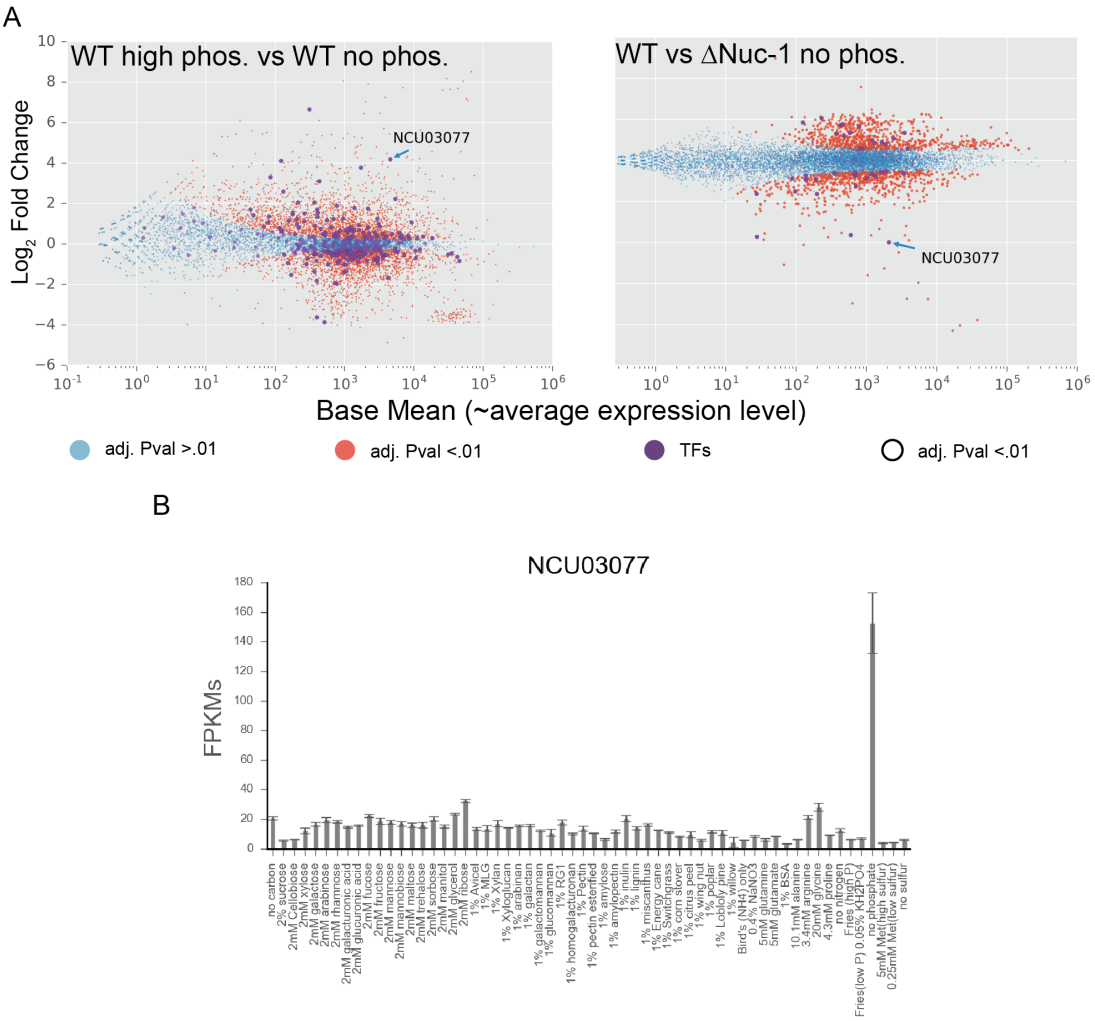


Figure 4-15 Identification of NCU03077, a TF induced under phosphate starvation conditions. *A) Differential expression analysis between WT high phosphate vs WT no phosphate and WT vs Δnuc-1 exposed to no phosphate, where the first of the two conditions are the reference conditions. Average expression of each gene is plotted along the x axis while log_2 fold change is plotted on Y. Significantly differentially expressed genes with p adjusted values below 0.01 are represented in red, whereas statistically insignificant genes are represented in blue. Further colors denote special groups of genes of interest, and outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.01). B) Normalized FPKMs for NCU03077 across all WT conditions. Error bars represent standard deviation from biological triplicates.*

NCU03077 is a previously un-described transcription factor with a basic Helix loop helix DNA binding domain. The closest blast hit to a known protein is a murine sterol regulatory element binding protein, with 20% identical and 40% positive amino acid identity. To further understand the role of NCU03077 in *N. crassa*, we performed RNAseq on ΔNCU03077 strain under phosphate starvation conditions. When we compared this data to WT data, we observed very few genes

differentially expressed above 4 fold between the two conditions (Figure 4-16). These data indicate that NCU03077 plays an inconsequential role in phosphate starvation response at 4 hours. It is possible that NCU03077 may have a greater role at a later time point. The few genes that are differentially expressed at 4 hours include a few sugar transporters and genes encoding CAZymes. This may indicate a role in cross talk between carbon and phosphate metabolism, however we are unable to draw any strong conclusions from these data currently. Additionally we see up regulation of NCU03078 a gene encoding a hypothetical protein adjacent to NCU03077. This increase in transcript abundance may be due to the presence of the NCU03077 promoter acting on the downstream NCU03078 in the ΔNCU03077 strain.

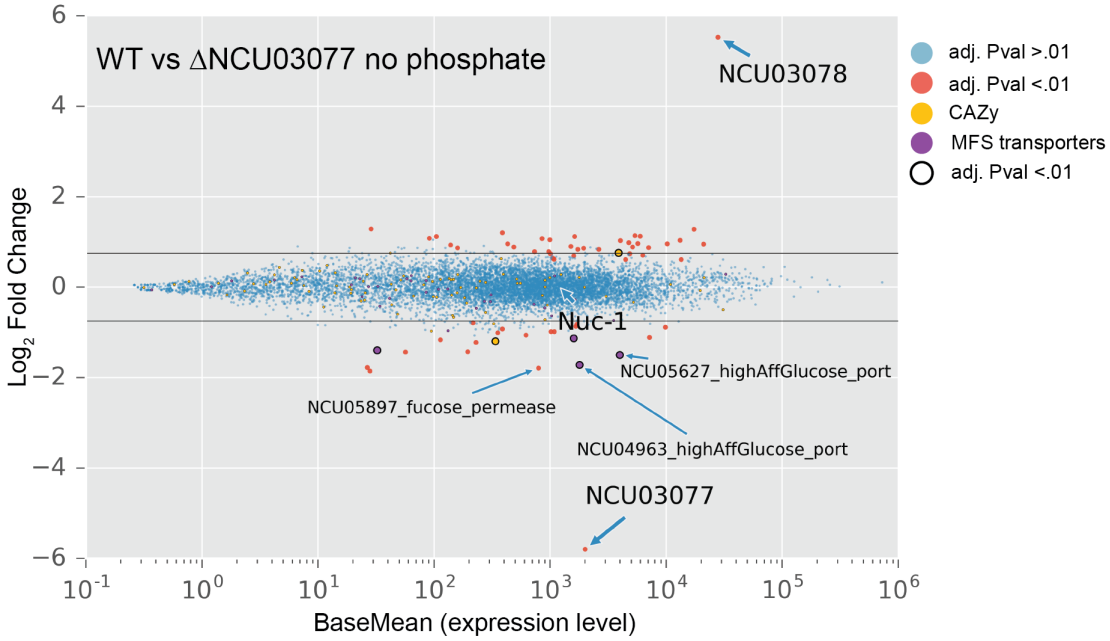


Figure 4-16 Differential expression of genes in WT vs ΔNCU03077 under no phosphate conditions. *Average expression of each gene is plotted along the x axis while log_2 fold change is plotted on Y. Significantly differentially expressed genes with P adjusted values below 0.01 are represented in red, whereas statistically insignificant genes are represented in blue. Further colors denote special groups of genes of interest, and outlined circles represent those genes that are significantly differentially expressed (adj. Pval. < 0.01). Horizontal lines added to the scatter plot represent + and – .75 log_2 fold change.*

4.6.3. ***Discussion***

We were able to replicate findings for the basic function of the NUC-1 transcription factor. During phosphate starvation, an active NUC-1 protein promotes expression of key genes involved in phosphate acquisition. These include a number of phosphatases, inorganic phosphate transporters and nucleases, some

previously unmapped. We show that in the *Δnuc-1* strain, expression of a large number of additional genes are both positively and negatively affected. This indicates that NUC-1 either directly or in directly regulates additional processes to phosphate acquisition. NUC-1 appears to repress expression of a number of ribosomal proteins and RNA processing genes, indicating its role in dampening protein biosynthesis during phosphate starvation. Additionally, our data shows that it is positively regulating not only phosphate acquisition but also a small set of carbon metabolism genes. To note, there are two transcription factors, *xlr-1* and *hem-1*, both positive regulators of hemicellulases, where their expression is reduced when *nuc-1* is deleted.

Without phosphate, the cell is unable to perform many basic functions and it makes sense that Nuc-1 mediates many aspects of phosphate starvation response aside from phosphate acquisition. Nuc-1 is at the end of direct signal transduction pathway for the detection of phosphate (see section 3.6.1). Logistically, it is not surprising that this transcription factor also regulates other crucial cellular responses to scarcity of phosphate, as exemplified by dampening of protein biosynthesis and de-repression of PCWDE inducing TFs.

Chapter 5. **DAPseq reveals complex regulatory mechanisms for control of PCWDEs**

5.1. **Introduction**

DNA affinity purification sequencing (DAPseq) is a method used to assay direct binding sites of transcription factors to the genome (179). The most similar technology to DAPseq is ChIPseq, an acronym of Chromatin Immuno-Precipitation Sequencing. In ChIPseq, cells with a epitope-tagged TF are fixed and lysed. The resulting lysate is sonicated to shear DNA. An immunoprecipitation with antibodies specific for the epitope-tag is performed to pull down tagged protein bound to sheared DNA. Finally, the DNA is eluted from protein and amplified for sequencing. This technique produces binding data that includes all sites that the target TF binds with all of the cofactors present *in vivo*. DAPseq is a similar technique that captures binding sites of TFs, but using a simpler and more high throughput method. Epitope-tagged TFs are transcribed and translated from plasmids *in vitro*, combined with a genomic DNA library prep, incubated and washed to remove unbound DNA. DNA is eluted from the tagged TF, amplified and sequenced (179). The resulting data can be different from ChIP data because cofactors present *in vivo* may not be present *in vitro* and TFs have to bind to DNA independently of chromatin modifications or other protein factors. However, this method is technically easier, because no transformation or expression validation is needed. Thus, limited troubleshooting is required in comparison to ChIPseq.

We used DAPseq to assay the direct binding targets of a number of TFs, in order to understand whether expression changes observed in our RNAseq datasets are due to direct binding of target TFs and activation of target genes or due to downstream affects of a TF deletion. We assayed 23 TFs. These include the TFs we

found of interest, as well as previously described and characterized TFs that served as controls. These previously characterized regulators allowed us to evaluate whether the DAPseq protocol was working and thus results could be extended for the other untested fungal TFs. Below is a table of the TFs we conducted DAPseq on. Highlighted in red are TFs from which we were able to call a substantial number of peaks (>~100 peaks). Data from remaining TFs yielded no callable peaks or a low number of biologically uninformative peaks.

Table 5-1 Transcription factors assayed for direct binding by DAPseq

Gene ID	TF name and description
NCU08294	*nit-4*
NCU04731	*sah-2*
NCU04731_Trunc	*sah2 truncated*
NCU04295	*pdr-2*
NCU09033	*pdr-1*
NCU00445	*man-1*
NCU08807	*cre-1*
NCU01074	*hem-1*
NCU03077	phosphate starvation TF
NCU03536	*cys-3*
NCU02142	lignin up regulated TF
NCU07788	*col-26*
NCU07705	*clr-1*
NCU06971	*xlr-1*
NCU06971A828V	*xlr-1* constitutive mutant
NCU00340	*pp1*
NCU07392	*adv1*
NCU05414	*ara-1*
NCU08042	*clr-2*
NCU09068	*nit-2*
NCU09315	*nuc-1*
NCU03725	*vib-1*
NCU01154	*sub-1*

We obtained binding results via DAPseq for a seven TFs so far (NIT-2, CYS-3, ADV-1, CLR-2, XLR-2, CRE-1, VIB-1). TFs XLR-1, CLR-1, CLR-2 and ADV-1 have published ChIPseq data that we use to validate the success of our DAPseq data sets. We show that DAPseq is robust technique that can be used to assay direct binding of fungal transcription factors. This data greatly expands our understanding of the role of previous studied nitrogen and sulfur metabolism TFs NIT-2 and CYS-3. Our data regarding two conserved TFs, CRE-1 and VIB-1, reveal new information

regarding how CCR and nutrient acquisition occur in *N. crassa* that are likely relevant to the rest of the fungal kingdom.

5.2. Materials and Methods

Genomic DNA library prep

For genomic DNA isolation, the FGSC 2489 strain was grown on liquid Vogel's minimal media (VMM) for 24 hours at 25˚C. Mycelia was filtered using Whatman #1 filter papers, and collected into 2ml tubes for flash freezing in liquid N_2 and cell rupturing. Cell rupturing was conducted by adding 1mm silica beads and along with DNA lysis buffer (0.05M NaOH, 1mM EDTA, 1% TritonX) and placed into bead beater for 1 minute. DNA was purified using DNeasy Blood & Tissue kit (Qiagen Inc.). DNA was sheared to 300BP peak using Covaris LE220 sonicator. Size selection for sheared DNA was performed using AMPure XP beads to remove DNA above and below target molecular weight. Initially, sheared DNA was mixed in with AMPure XP beads (in PEG-8000) at a ratio of 100:60. At this ratio, beads bind DNA with molecular weight above 700BP. Supernatant from this primary binding was taken and added to new beads where final ratio of DNA solution to PEG-8000 was at 100:90. At this ratio, DNA below ~300 do not bind to AMPure XP beads, and remaining DNA can be eluted for library preparation. KAPA library kit for illumine sequencing was used to prepare final libraries and stored at -20˚C for later use.

Cloning for TF expression plasmid

TF open reading frames were amplified from cDNA generated using RNA to cDNA ecodry premix (Clonetech). Amplified TFs sequences were inserted into expression vector containing T7 and SP6 promoters upstream of HALO tag. Vector backbone was amplified and assembled with TF ORFs using Gibson assembly and transformed into competent *E. coli* cells for storage and production.

Transcription, translation and DNA affinity purification (DAP)

In vitro transcription and translation of TFs was performed using Promega TnT T7 Quick Coupled Transcription/Translation System by incubating 1ug of plasmid DNA with 60ul of TnT Master Mix and 1.5ul of 1mM methionine. Expression of TFs was verified using Western blot analysis with Promega Anti-HaloTag monoclonal antibody. DAP was performed by incubating 20ng of genomic DNA libraries along with 1ug salmon sperm with completed TnT reactions and Promega Magne HaloTag Beads for one hour at room temperature. Bead bound proteins along with protein bound DNA were washed with 2.5% Tween20 in PBS and eluted from beads with high temperature elution at 95˚C for 5 minutes. Eluted DNA was amplified for final libraries using KAPA Hifi polymerase for 12-16 cycles of PCR. A graphic of the full DAPseq protocol from O'Malley et al. 2016 is presented below (179).

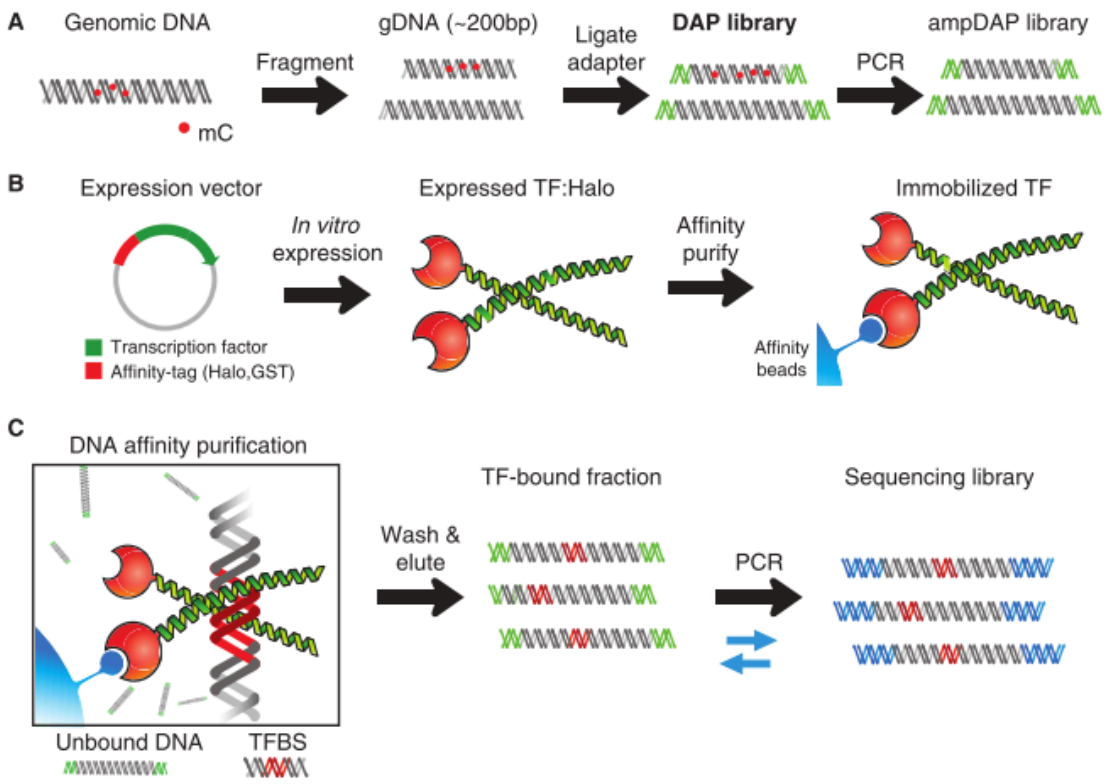


Figure 5-1 DAPseq protocol schematic. *Protocol from O'Malley et al. 2016* (179)*, used to conduct DAPseq on N. crassa TFs.*

Analysis

Filtered reads were mapped against *N. crassa* OR74A genome (v12) using bowtie2 v2.3.2 (114). Samtools (180) was used to convert .sam to .bam files and to create .bai index files for viewing reads on IGV (Integrated Genomics Viewer). MACS2 (181) with pvalue cutoff at 0.001 was used for calling peaks. A negative control data set consisting of DAP pulldown with promega TnT master mix with no plasmid added, salmon sperm and genomic DNA was also input into MACS2 as the control condition.

Motif construction

To construct motifs for transcription factors, available RNAseq data was used to eliminate false positive or biologically uninformative DAPseq peaks. This generally involved selecting genes differentially expressed between WT and transcription factor deletion strains, then further eliminating genes with no direct binding evidence. Sequences of true positive peaks were collected using a custom python script that reads in genomic position each peak from MACS2 output file, and ascertains the sequence from *N. crassa* FGSC 2489 genomic DNA FASTA file. The script output is a fasta file with sequences from all true positive peaks. Output FASTA files were input into MEME or DREME v4.12.0 with flags maxw =20, minsites = 5, nmotifs = 8, denoting max width of motif, minimum number of sites for each motif and number of motifs to generate respectively.

5.3. DAPseq validation using CLR-2 and XLR-1 in comparison to ChIPseq data

5.3.1. *Introduction*

DAPseq is an untested method for fungal transcription factors. We wanted to ensure that the DAP protocol would be successful for fungal TFs. The *N. crassa* TFs we were interested in tended to be much larger than the plant TFs where DAPseq has been successful in the past (179). Three transcription factors that were ideal for these validation experiments were XLR-1, CLR-1 and CLR-2. Unfortunately we have been unable to enrich for any DNA in the CLR-1 DAP. We were able to successfully conduct DAPseq on XLR-1 and CLR-2.

ChIPseq has been performed with XLR-1 and CLR-2 (182) to assay direct binding of these two TFs in an *in vivo* setting. The authors used a mutant XLR-1^{A828V} strain that remains active under glucose conditions (182) to assay XLR-1 target genes by RNAseq. By combining RNAseq results with XLR-1-GFP ChIPseq data, a target list of promoter regions of genes that were both bound by XLR-1 and up regulated in the XLR-1^{A828V} strain was identified. This list includes 23 genes mostly composed of xylan/xylose metabolic genes (182). A similar set of genes was compiled by taking the overlapping set of genes that show up regulation in a *clr-2* over-expression strain and also had binding of CLR-2-mCherry in their promoter regions. This set includes 54 genes composed mostly of cellulases, LPMOs and hemicellulases. By taking overlapping gene sets between RNAseq and ChIPseq data, the authors were able to filter out a large number of likely false positive peaks.

We conducted DAPseq using halo tagged XLR-1^{A828V} and CLR-2. We combined these data sets with RNAseq of Δ*xlr-1* and Δ*clr-2* compared to WT cells on xylan and Avicel, respectively. We counted genes as down regulated if their expression was reduced 2.25 fold change or more as calculated by DEseq (117) with adjusted Pvalue less than 0.01. Genes in this list that were also bound by the TF according to our DAPseq dataset we concluded are directly regulated by the TF.

5.3.2. *Results*

Exploring XLR-1 DAPseq targets

We found 63 genes directly regulated by XLR-1. These genes are both reduced in expression in the *Δxlr-1* strain and are bound by XLR-1 according to DAPseq. This is out of 338 genes in total bound in at least one position within 3 kb upstream of the ATG start site. This is a 18.6% true positive rate. We also observed that the genes most differentially expressed and bound are bound within 1kb of the ATG start site (Figure 5-2A). This indicates that binding sites within 1kb are more likely to be true positives, while genes 2kb or 3kb away are more likely to be false positives. We observe fewer differentially expressed genes bound between 1kb and 3kb from the presumptive ATG translation start site as well. We further observed that as the differential expression a gene increases the more likely it is bound by XLR-1. Alternatively, binding rate remains the same for genes down regulated in

WT cell as compared to Δ*xlr-1* cells (Figure 5-2B). These observations confirm the role of XLR-1 as a positive regulator rather than a negative regulator. Positive regulation by XLR-1 is further supported by comparatively low number of genes that show greater expression in the Δ*xlr-1* strain than in the WT strain.

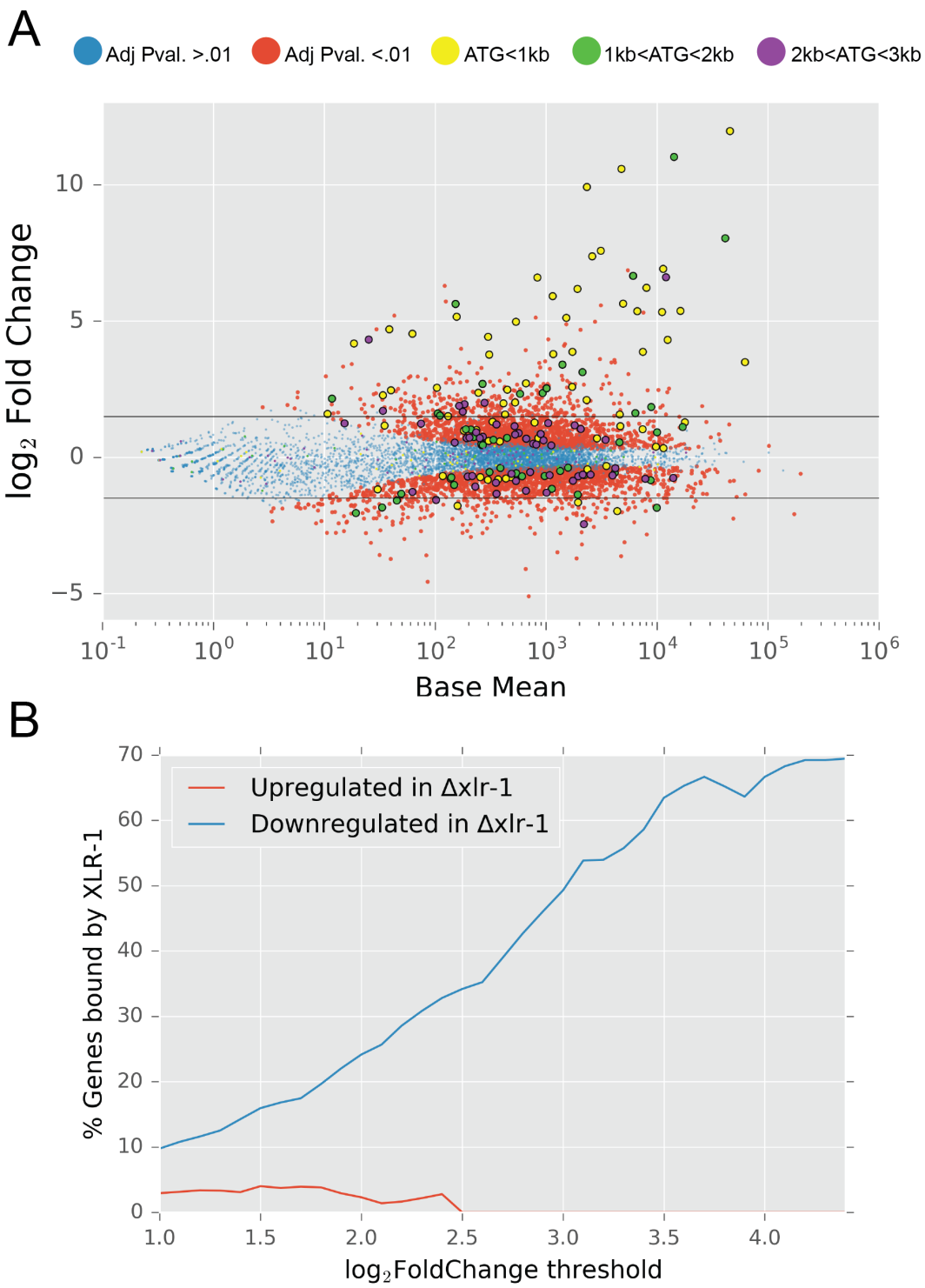


Figure 5-2 RNAseq and DAPseq provide direct targets of XLR-1. *A) plot of differential expression of genes between Δxlr-1 and WT, where Δxlr-1 is the reference strain. XLR-1 DAP bound genes are highlighted in green, yellow and purple, with colors denoting distance of binding peak from ATG start site. Black outlined data points indicates differential expression p adjusted value < .01 B) line plot of % of genes bound by XLR-1 as a function of log_2 fold change threshold. Two lines are plotted, one*

representing down regulated genes in Δxlr-1 (blue) and the other up regulated genes in Δxlr-1 (red). Positive slope of down regulated genes indicate XLR-1 is a positive regulator.

XLR-1 DAPseq comparison to ChIPseq

As stated in the introduction to this chapter, ChIPseq of XLR-1 in combination to RNAseq of expression profile of the *xlr-1*^{A828V} mutant yielded a set of 23 genes considered directly regulated by XLR-1 (31). We compared this set to our set of 63 genes that we conclude are directly regulated by XLR-1. We observed that 20 genes identified by ChIP data overlapped with our genes identified by DAPseq. The three genes that are only in the ChIP data are a β-xylosidase (NCU00709), an acetyl xylan esterase (NCU05159), and a hypothetical protein (NCU06490). We were surprised that β-xylosidase (NCU00709) and acetyl xylan esterase (NCU05159) were not present in our dataset. These two genes are two of the highest differentially expressed genes in *Δxlr-1* mutant compared to WT cells. However, we did observe an increased number of reads in β-xylosidase (NCU00709) promoter region that did not reach our peak-calling cutoff. We also observed a large peak about ~6KB away from the ATG start site of NCU05159, where NCU05159 is also the closest gene.

Of the 63 gene set, 43 genes were in the DAPseq set only. These 43 genes include a number of xylanases we expect to be under direct regulation by XLR-1. These include endoxylanases NCU02855 and NCU05924, acetyl xylan esterases NCU00710 and NCU03010, and β-xylosidase NCU009923. The rest of the set contains several arabinases, pentose catabolic genes, sugar transporters but also a number of hypothetical proteins and other xylan/xylose unrelated genes. Overall, many of the DAPseq only genes make logical sense for being directly regulated by XLR-1. Both techniques seem to be missing certain genes that have high probability of being directly regulated by XLR-1.

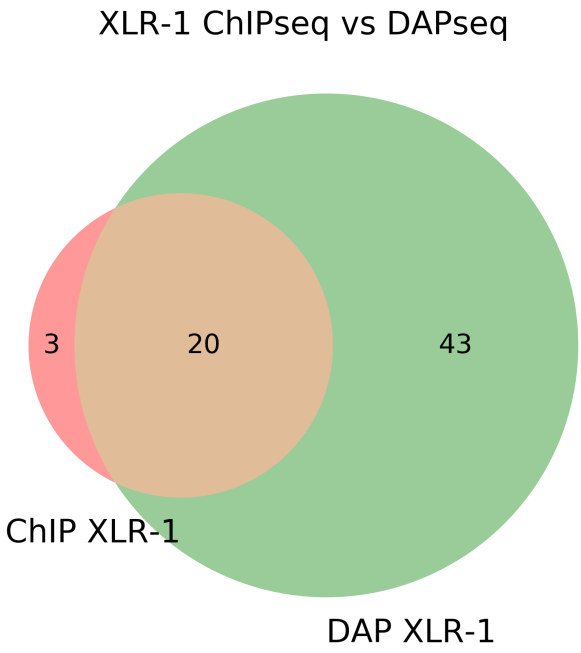


Figure 5-3 Venn diagram of XLR-1 directly regulated genes: ChIPseq vs DAPseq. *Numbers represent the number of genes directly bound in each data set. DAP bound genes must show greater than 2.25x fold decrease in Δxlr-1 strain compared to WT, and must also have promoters bound within 3kb upstream of the ATG start site.*

We used sequences of the direct binding peaks of XLR-1 to determine a binding motif. We found a single significant motif with an E-value of $3x10^{-23}$. This motif is almost identical to the motif determined by ChIPseq in Craig et al. 2015 (31).

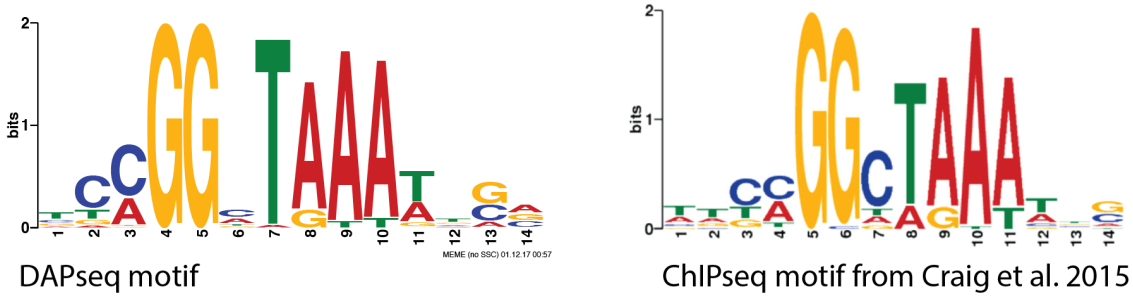


Figure 5-4 XLR-1 DAPseq vs ChIPseq binding motif

Exploring CLR-2 DAPseq targets

We identified 88 genes that are 2.25x down regulated in the *Δclr-2* mutant and also bound by CLR-2 according to DAPseq. This set was filtered down from the whopping 1683 total genes with promoters bound up to 3 kb upstream of the ATG

in the DAPseq dataset. CLR-2 was one of the genes with the most binding sites out of any of the TFs we tested. This result could be explained by a strong affinity to DNA, but may also be due to other unknown stochastic variables. Similar to XLR-1 bound genes, down regulated genes in *Δclr-2* strain were more likely to have binding sites closer to the translational start site (ATG). Genes bound within 1kb of the ATG start are more abundant within the set of *Δclr-2* differentially expressed genes. However, many genes that we would expect to be directly regulated by CLR-2 have promoter-binding peaks further away (Figure 5-5A). Like XLR-1, we expected CLR-2 to show characteristics of a positive regulator (27, 31). This expectation was supported by a positive slope in the plot of percentage of genes bound as a function of increasing $\log_2$ fold change threshold for genes down regulated in the *Δclr-2* mutant (Figure 5-5B). Interestingly, the slope evens out to zero as above 2.5 $\log_2$ fold change. This result is surprising because we expect the most down regulated genes in the *Δclr-2* mutant to be bound by CLR-2. It is possible that these genes require additional in vivo factors for binding to these promoters.

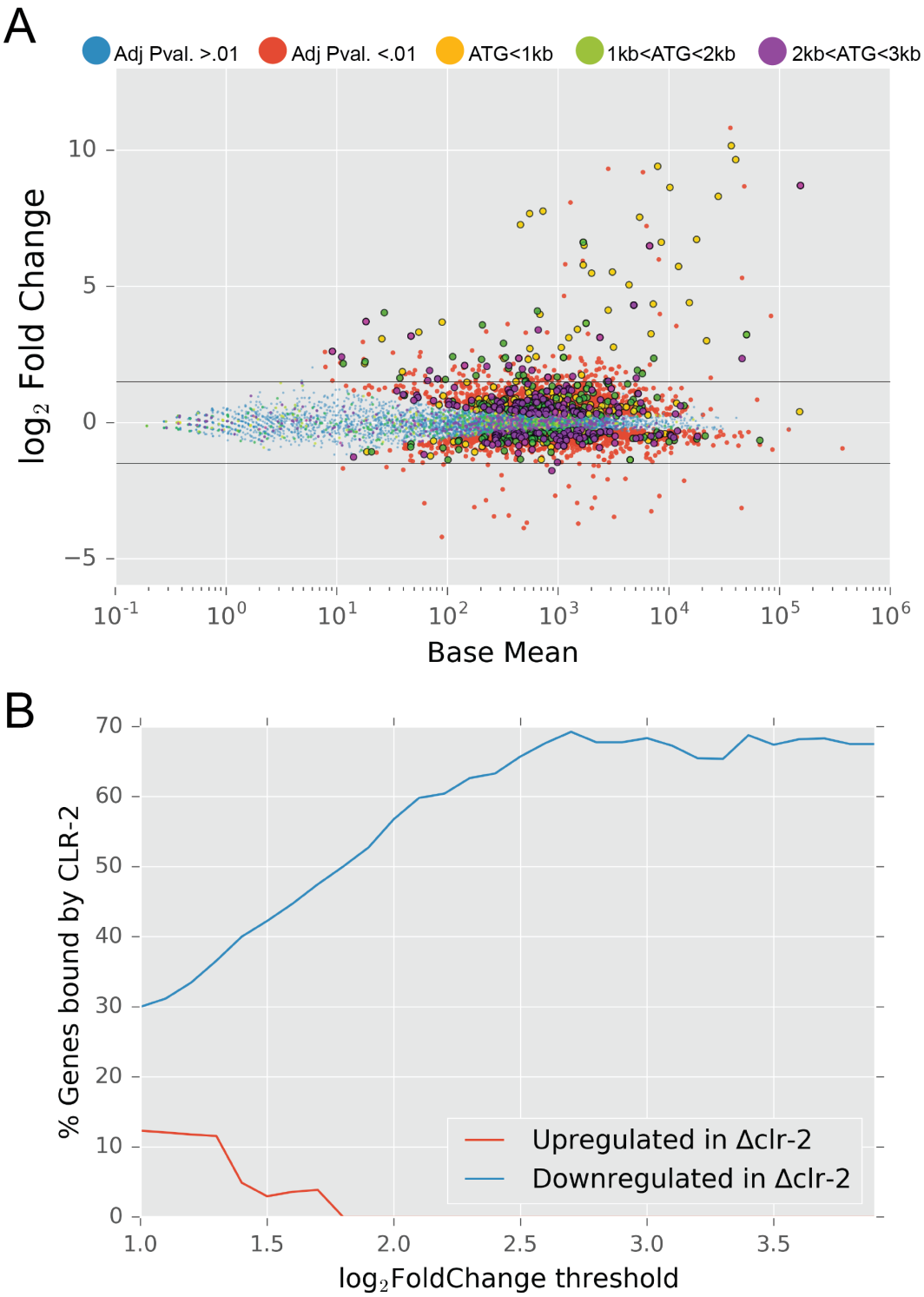


Figure 5-5 RNAseq and DAPseq provide direct targets of CLR-2

A) plot of differential expression of genes between Δclr-2 and WT, where Δclr-2 is the reference strain. CLR-2 DAP bound genes are highlighted in green, yellow and purple, with colors denoting distance of binding peak from ATG start site. Black outlined data points indicates differential expression p adjusted value < .01 B) line plot of % of genes

bound by CLR-2 as a function of log_2 fold change threshold. Two lines are plotted, one representing down regulated genes in Δclr-2 (blue) and the other up regulated genes in Δclr-2 (red). Positive slope of down regulated genes indicate CLR-2 is a positive regulator.

Comparison of CLR-2 DAPseq to ChIPseq

As stated in this sections introduction, CLR-2 ChIPseq in combination with RNAseq of a CLR-2 mis-expression strain under *ccg-1* promoter yielded a set of 54 genes that are directly regulated by the TF (31). We compared this set of genes to our set of 88 genes. We found that 34 genes were overlapping between the two data sets but significant numbers of genes were unique to both. The set of ChIPseq genes includes a number of PCWDEs that we would expect to be bound by CLR-2 such a xylanases (NCU05924, NCU07725), glucanases (NCU09680, NCU5121), an LPMO (NCU09764) and the cellodextrin transporter CDT-2 (NCU08114). For many of these genes, promoters have an increased number of reads that did not pass the peak calling threshold. Our CLR-2 data is especially noisy in comparison to other DAPseq datasets. It is possible that these are real peaks and CLR-2 binding sites, but not called due to the high frequency of seemingly random peaks associated with the data. It is also possible that CLR-2 may bind some of these genes better with *in vivo* co-factors that are not present in DAPseq dataset.

CDT-2 is peculiar in that it is up regulated in the CLR-2 mis-expression strain and bound also by CLR-2 in the ChIPseq dataset (31), although it is not nearly as strongly regulated by CLR-2 as many of the cellulases. However, our data shows that expression of CDT-2 does not change very much (less than 2X fold change) in a *Δclr-2* mutant (Table 5-2). We also did not find any binding peaks or possible uncalled peaks in the CDT-2 promoter. We also observe that *cdt-2 is* positively regulated strongly by both CLR-1 and XLR-1 displayed by RNAseq differential expression analysis shown in the table below (Table 5-2). These data indicate that CLR-2 is likely a minor regulator of *cdt-2*, which can explain the lack of binding in the DAPseq data.

Table 5-2 Log_2 fold change of *cdt-2* in *Δclr-2, Δclr-1* and *Δxlr-1* versus WT

WT vs *Δclr-2* 1% Avicel	WT vs *Δclr-1* 1% Avicel	WT vs *Δxlr-1* 1% xylan
-0.84	-3.01	-6.61

We also did not observe *clr-3* (NCU05846) in our DAPseq data set. CLR-3 is a repressor of CLR-1 activation (183) and is part of the ChIPseq set of directly regulated genes. Huberman et al. 2017 reported that CLR-3 represses CLR-1 in the absence of Avicel (183). When we take a closer look at our data, we see a strong peak in the promoter of *clr-3* ~1.5KB upstream of the ATG with 13-fold enrichment of reads. However, in our RNAseq dataset, *clr-3* is only differentially expressed at 1.93 fold, which is below our cutoff of 2.25 fold change. Low differential expression might be due to our comparison of WT to a *Δclr-2* strain rather than the *clr-2* mis-expression strain in the ChIPseq data set. Indeed, expression of *clr-3* is

approximately 6X up regulated in the *clr-2* mis-expression strain under no carbon compared to WT (31).

We observed 53 genes present only in the CLR-2 DAPseq data set. This set includes a number of genes we would expect to be in the CLR-2 regulon such as glucanase (NCU05104), β-glucosidases (NCU04952, NCU07487), rhamnogalacturonan lyase (NCU05598) and acetyl xylan esterase (NCU00710). A large proportion of the remaining genes encode hypothetical proteins. Overall, the ChIPseq-DAPseq overlapping dataset contains the majority of cellulase genes. However, both data sets have seemed to miss key genes predicted to be regulated by CLR-2, perhaps because cutoffs are too stringent for this particular data set.

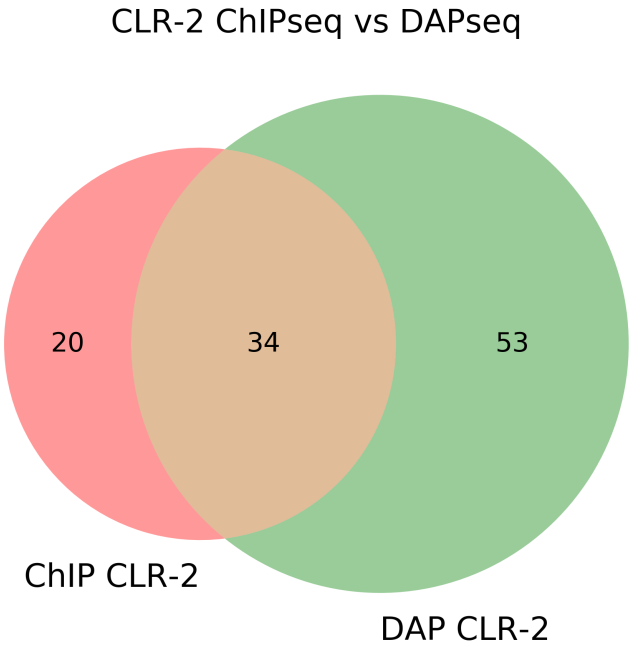


Figure 5-6 Venn diagram of ChIPseq vs DAPseq direct targets of CLR-2.
Numbers represent the number of genes in each category.

We used sequences of the direct binding peaks of CLR-2 to determine a binding motif. We found a single significant motif with an E-value of 1.2×10^{-13}. This motif is very similar to the motif determined by ChIPseq in Craig et al. 2015 (31).

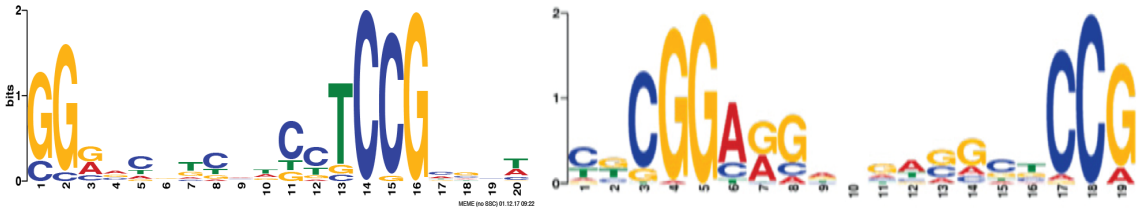


Figure 5-7 CLR-2 DAPseq vs ChIPseq binding motif

5.3.3. *Discussion*

DAPseq results were consistent with results observed from ChIPseq data from Craig et al. 2015 (31). Overall, the XLR-1 DAPseq data set was much cleaner than CLR-2 data set in terms of ratio of true peaks to false peaks. The XLR-1 DAPseq data set also picked up the majority of genes suspected to be directly regulated by XLR-1 from RNAseq, including all but three of the ChIP gene set. CLR-2 data was noisier in comparison. Many peaks were found in promoters of genes unrelated to cellulose metabolism or even plant cell wall metabolism. These genes were also not down regulated in *Δclr-2*. For these reasons, we conclude that these genes are either false positive binding sites or biologically irrelevant binding sites.

When compared to ChIPseq results, the majority of ChIP gene set was also shown to be directly regulated in the DAP set. In general, the final sets for DAPseq were a bit larger. This is likely due to different cutoffs used for the RNAseq. RNAseq datasets were also fundamentally different as the ChIPseq experiments compare TF mis-expression to WT, while our RNAseq compares TF deletion strains to WT. Thus, results are not directly analogous. All in all, both ChIPseq and DAPseq were successful in finding the core regulons of XLR-1 and CLR-2. DAP XLR-1 was arguably more successful than ChIP XLR-1, while both DAP CLR-2 and ChIP CLR-2 are likely incomplete data sets.

5.4. MAN-1 is a positive regulator of amino acid catabolism as shown by DAPseq

5.4.1. *Introduction*

Man-1 (NCU00445) is one of the previously undescribed transcription factors we explored in Chapter 4. Its transcription is significantly upregulated under 2mM mannose in comparison to our other WT conditions. RNAseq of the Δ*man-1* strain on 2mM mannose showed that deletion of the transcription factor results in a significant number of differentially expressed genes in comparison to WT both upregulated and downregulated (see section 4.4.2). We hypothesized that part role of MAN-1 is to directly promote expression of a small number of PCWDEs that we observe are downregulated in the *Δman-1* strain. We further hypothesized that the role of MAN-1 was related to carbon metabolism, due to its uniquely high expression under 2mM mannose. However, DAPseq revealed that MAN-1 directly binds to a number of amino acid catabolic genes, specifically ones critical for branched chain amino acid and proline catabolism.

5.4.2. *Results*

DAPseq results for halo-tagged MAN-1 revealed a number of direct binding sites within amino acid catabolism related gene promoters that are supported by RNAseq differential expression data between WT and Δ*man-1* strains. Overall we found 291 unique genes bound within 3000bp upstream of the ATG start site. 6 of these genes had a secondary binding site within the 3000bp promoter region. We highlighted the MAN-1 bound genes in a plot of differential expression data from RNAseq. We found that some of the most downregulated genes in the *Δman-1*

mutant are also directly bound by the MAN-1 (Figure 5-8A). Furthermore when we plot the percentage of genes bound as a function of log_2 foldchange threshold, we see that only genes downregulated in *Δman-1* become increasingly enriched with greater log_2 foldchange scores (Figure 5-8B). These data indicate that MAN-1 is likely a positive regulator. We therefore conclude that the genes that are downregulated in *Δman-1* and bound by DAPseq are likely direct targets of MAN-1 for positive regulation. We use an arbitrary 4x fold change (log_2 foldchange > 2) threshold to include in this list. These genes are listed in Table 5-3

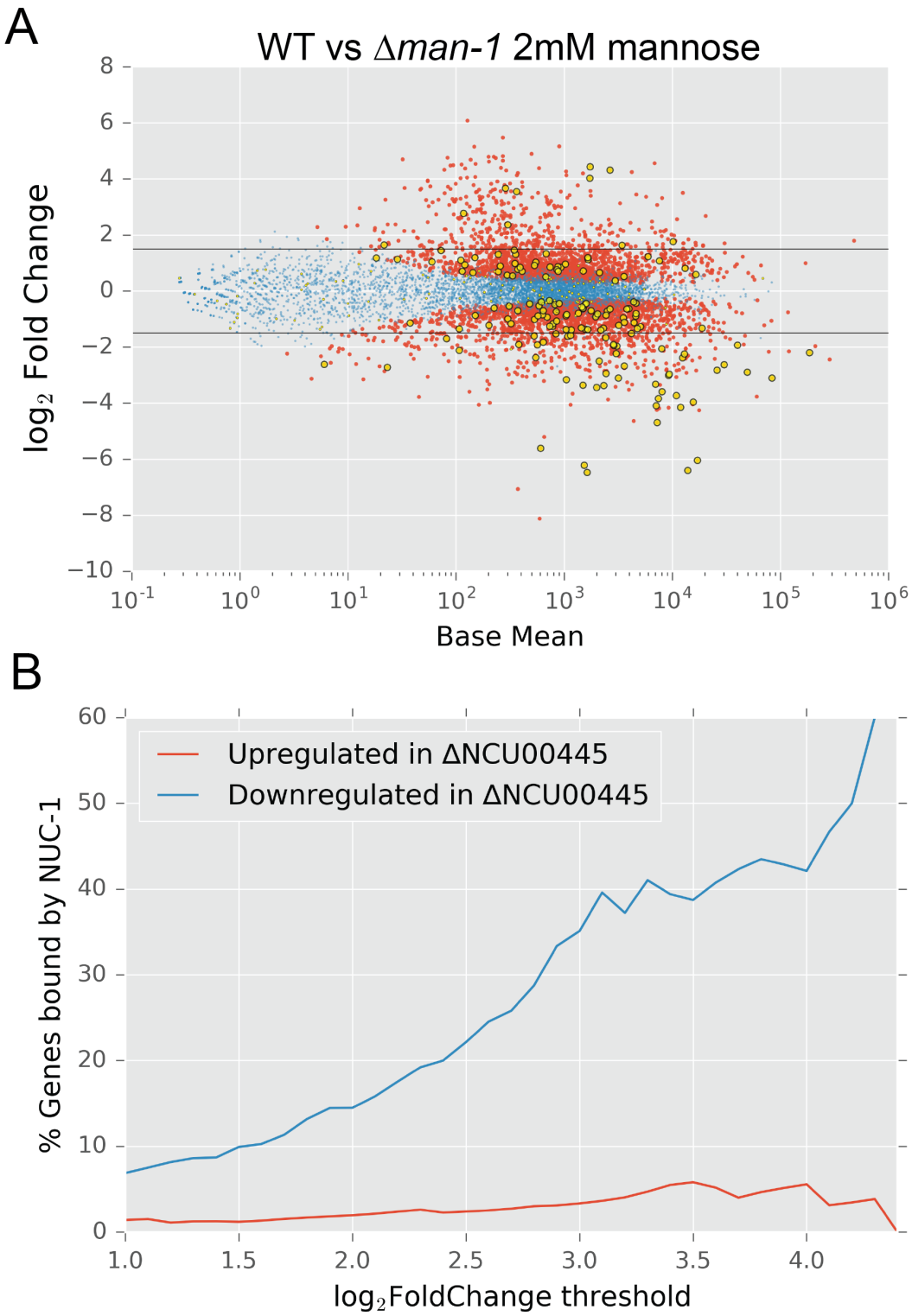


Figure 5-8 RNAseq and DAPseq provide direct targets for MAN-1 *A) Plot of differential expression of genes between Δman-1 and WT, where Δman-1 is the reference strain. MAN-1 DAP bound genes are highlighted in green, yellow and purple, with colors denoting distance of binding peak from ATG start site. Black outlined data*

points indicates differential expression p adjusted value < .01 B) line plot of % of genes bound by MAN-1 as a function of log_2 fold change threshold. Two lines are plotted, one representing down regulated genes in Δman-1 (blue) and the other up regulated genes in Δman-1 (red). Positive slope of down regulated genes indicate MAN-1 is a positive regulator.

Table 5-3 Genes directly bound by MAN-1

Gene ID	Annotation	fold enr.	Dist. ATG
NCU00445	hypothetical protein	5.39	96
NCU02704	branched-chain alpha-keto acid dehydrogenase E2 component	7.78	300
NCU01829	hypothetical protein	6.97	556
NCU06895	cytochrome P450 4A5	7.67	304
NCU06618	metallopeptidase MepB	6.64	672
NCU02936	proline oxidase	3.90	451
NCU00721	proline-specific permease	5.99	356
NCU02126	isovaleryl-CoA dehydrogenase	6.27	368
NCU03076	delta-1-pyrroline-5-carboxylate dehydrogenase	7.95	422
NCU05499	"homogentisate 1,2-dioxygenase"	6.37	331
NCU03913	2-oxoisovalerate dehydrogenase beta subunit	8.67	185
NCU02127	methylcrotonoyl-CoA carboxylase subunit beta	6.27	229
NCU02705	F1F0 ATP synthase assembly protein Atp10	7.78	369
NCU01830	4-hydroxyphenylpyruvate dioxygenase	6.97	807
NCU04266	hypothetical protein	8.32	741
NCU05537	fumarylacetoacetase	5.47	163
NCU17064	5-nitroimidazole antibiotic resistance protein	4.47	301
NCU08771	acetolactate synthase	5.65	1239
NCU08216	cystathionine beta-synthase	8.76	607
NCU07895	hypothetical protein	8.86	873
NCU09170	alpha-N-arabinofuranosidase II	4.28	561
NCU09345	"no message in thiamine-1, nmt-1"	6.61	600
NCU00591	methylcrotonoyl-CoA carboxylase subunit alpha	7.84	286
NCU05142	hypothetical protein	3.97	868
NCU08949	hypothetical protein	4.73	917
NCU09865	methylase	9.81	983
NCU08687	galactokinase	7.24	616
NCU08356	acetamidase	8.82	911
NCU08688	hypothetical protein	7.24	678
NCU09864	2-oxoisovalerate dehydrogenase alpha subunit	9.81	297
NCU06110	thiazole biosynthetic enzyme	7.54	1122

NCU02333	arginase	6.51	454

We used the binding peaks from this list of MAN-1 direct targets to identify a binding motif. By taking the sequences of the binding peaks, we used MEME v4.12.0 (184) to build a consensus motif common within all input sequences. A single motif with a significant E-value of $2.1x10^{-23}$ was discovered. The position of each motif was centered within each binding peak sequence, which is another indication that the motif is authentic (Data not shown). This motif is similar to motifs from other zn2cys6 TFs CLR-1 and CLR-2, both of which have two CCG or CGGs interspaced 5-11 base pairs between each other.

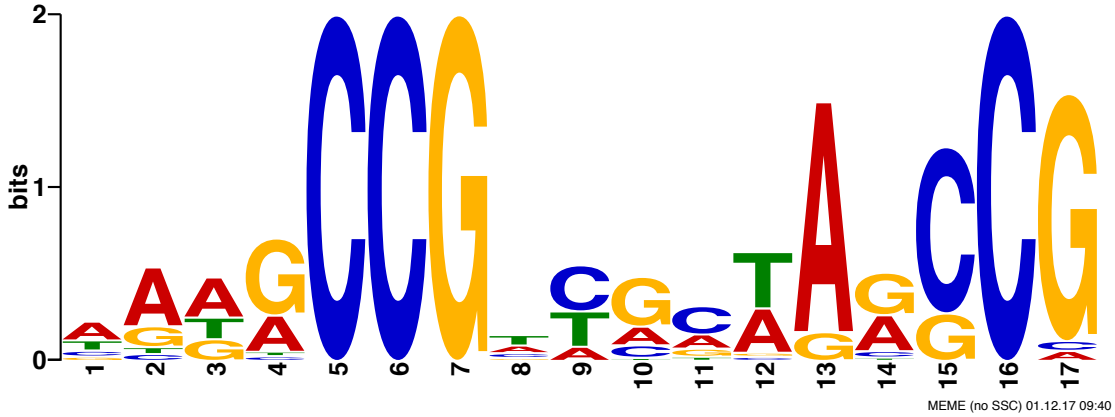


Figure 5-9 MAN-1 binding motif. E-value of $2.1x10^{-23}$

Closer examination of MAN-1 direct targets indicates a clear role of MAN-1 in the positive regulation of branched chain amino acid metabolism. We characterized the direct targets of MAN-1 with FunCat (121). We found that the top 15 functional categories enriched were amino acid metabolism related. Many of these functional categories are related to metabolism of branched chain amino acids valine, leucine and isoleucine. When we examined the direct target genes more closely, we noted seven genes involved in branched chain amino acid catabolism. These include acetoacetate synthase (NCU08771), methylcrotonyl-CoA carboxylase alpha (NCU00591) and beta (NCU02127) subunits, 4-hydroxyphenylpyruvate dioxygenase also known as KIC-dioxygenase (NCU05537), 2-oxoisovalerate dehydrogenase alpha (NCU09864) and beta (NCU03913) subunits and branched-chain alpha-keto acid dehydrogenase E2 component (NCU02704). These genes are highlighted in two alternate human leucine catabolism pathways below with the exception of 2-oxoisoalerate dehydrogenase, which is part of a third alternative catabolic pathway (Figure 5-10). Leucine and other isovaleric acids can be catabolized into acetyl-CoA that can then enter the citric acid cycle. The abundance enzymes required for branched chain amino acid metabolism that are direct targets of MAN-1 indicate that MAN-1 may be crucial for catabolism of these amino acids.

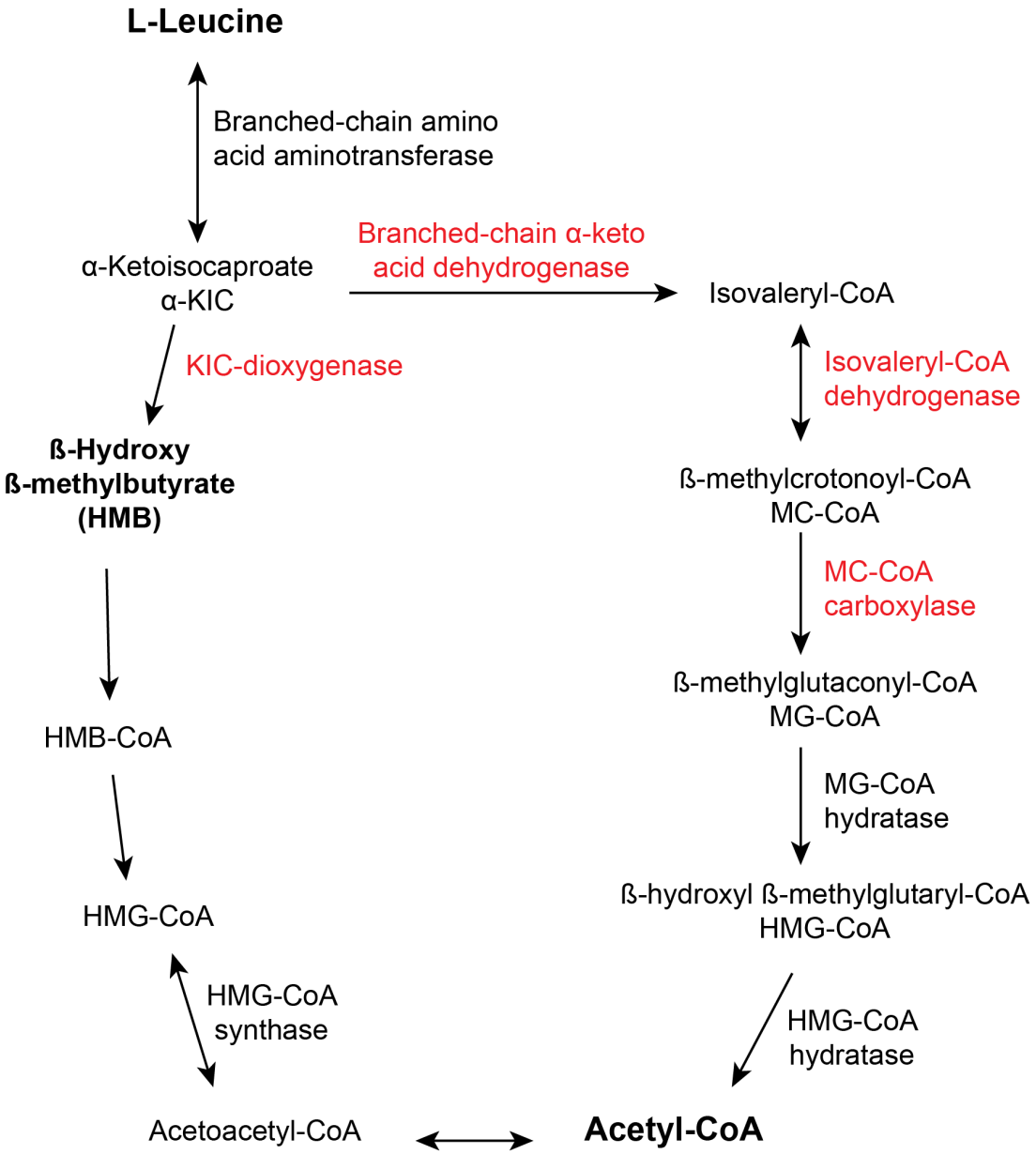


Figure 5-10 Human leucine catabolic pathway *Enzymes highlighted in red are downregulated more than 4 fold in the Δman-1 strain, and also bound by MAN-1 according to DAPseq. Leucine pathway adapted from information presented in Kohlmeier 2015* (185).

MAN-1 likely plays an additional roles in proline, phenylalanine and arginine catabolism. Two proline catabolic genes are present in the list of direct target genes of MAN-1. These are proline permease (NCU00721) and proline oxidase (NCU02936). Whether these are critical components of *N. crassa* proline catabolism is not yet known, but these two genes are the highest expressed annotated proline specific catabolic genes of four genes present in the Broad v12 annotation. The two additional genes are proline permease (NCU04468) and a proline iminopeptidase (NCU07415). According to DAP and RNAseq results, Proline permease (NCU04468)

is directly bound and under positive regulation by CYS-3, the canonical sulfur acquisition TF (See section 5.8). Three additional genes involved in phenylalanine and one arginase are direct targets of MAN-1. These are fumarylacetoacetase (NCU05537), homogentisate 1,2-dioxygenase (NCU05499) , 4-hydroxyphenylpyruvate dioxygenase (NCU01830) and arginase (NCU02333).

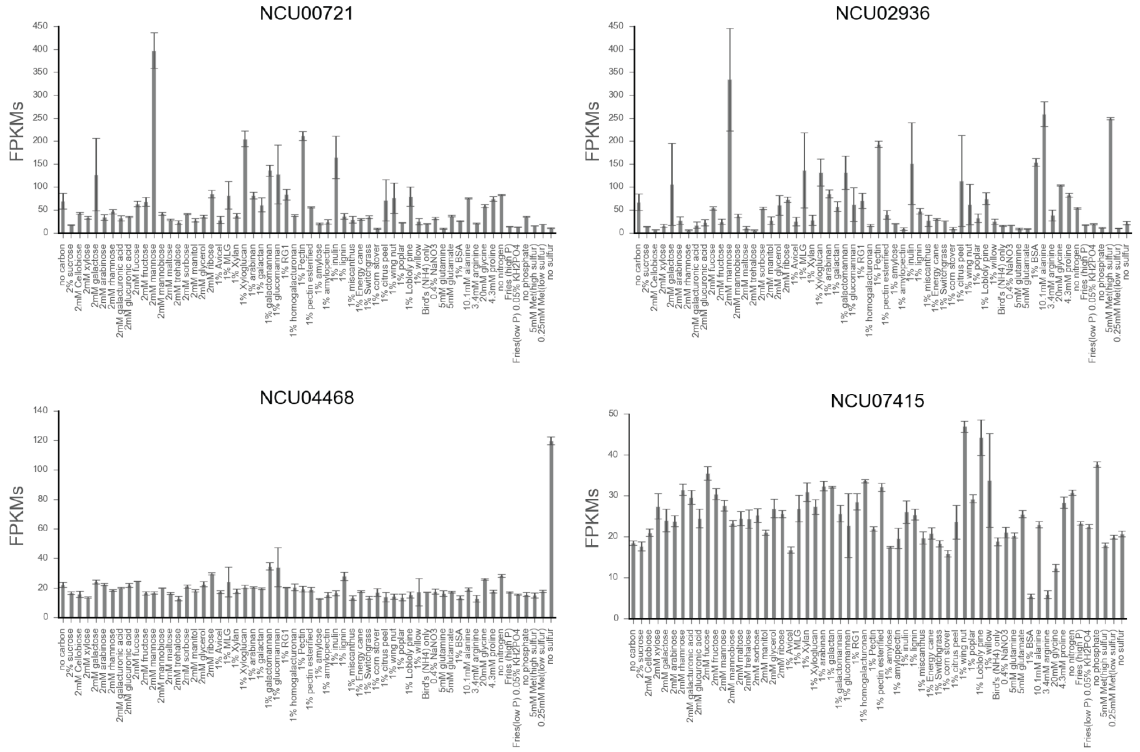


Figure 5-11 Proline catabolic gene expression profiles. *Expression as measured by FPKMs with error bars representing standard deviation from triplicate*

5.4.3. ***Summary and Discussion***

DAPseq of MAN-1 allowed us to identify the direct targets of the transcription factor. Direct targets were unclear from the RNAseq data of WT vs. *Δman-1* due to the high number of genes differentially expressed in both directions. By combining the two datasets we were able to determine a list of genes directly regulated by MAN-1. We utilized binding peaks of these genes to build a robust binding motif with high similarity to the binding motifs of CLR-1, CLR-2 and GAL4. The direct targets of MAN-1 are also highly enriched for catabolism of branched chain amino acids leucine, isoleucine and valine. Closer examination of the branched chain amino acid catabolism pathway indicates that MAN-1 is required for full expression of several critical genes for the pathway under 2mM mannose. We also found that MAN-1 is regulating a small number of proline, phenylalanine and arginine metabolic genes under 2mM mannose. At this time, we are uncertain of the connection between mannose and expression of MAN-1 along with amino acid catabolic genes. We expected direct binding partners of MAN-1 to be carbon related genes, however no such link was discovered. It is possible mannose is a signaling molecule for presence of glycosylated proteins that *N. crassa* can scavenge.

Mannose is abundant in N-linked glycosylation, which is abundant in secreted proteins in fungi and plants, and many other eukaryotic organisms (186). It is possible that presence of extracellular mannose is an indication that proteins are also present and available as a nutrient source. In response, *N. crassa* may upregulate amino acid catabolic genes such as those under positive regulation by MAN-1.

5.5. CRE-1 DAPseq makes us rethink the mechanism of CRE-1 mediated CCR

5.5.1. *Introduction*

CRE-1/CreA/MIG1 is a well-known repressor of PCWDEs and the major transcription factor in Ascomycota involved in carbon catabolite repression (CCR). Several studies have specifically examined the mechanism of action of CreA on its presumed target proteins in filamentous Ascomycota. Much of this work has been conducted in *A. nidulans* and explores interaction of CreA with xylanases and the transcription factor XlnR (*N. crassa* XLR-1). It was initially shown that CreA directly binds to the promoter region of a xylanase *xlnA* (187). Later it was shown that CRE-A also repressed expression of xylanases xlnB, but there was no evidence that it directly bound the xlnB promoter region (188). CreA also directly represses XlnR by experiments utilizing the *xlnR* promoter fused to a reporter gene in strains with either native or mutated *creA* (189). Thus xylanase expression is repressed through direct repression as well as repression of XlnR, a form of repression coined as the "double lock" mechanism (187).

A similar system of repression was hypothesized for CRE-1 in *N. crassa.* CRE-1 has been shown to affect the expression of a number of cellulases in *N. crassa* (27), and also to directly bind to the promoter regions of several PCWDEs and transporters through ChIP-PCR (190). These genes include glucose transporter (NCU04963), *col-26* (NCU07788), α-amylase (NCU09805), invertase (NCU04265), xylanase (NCU02885), hypothetical protein (NCU03181), *cbh-2* (NCU07190), xylanase (NCU07225) and *cbh-1* (NCU07340) (190). These genes were also shown to have the CRE-1/CreA binding motif 5'-SYGGRG-3' (191) in their promoter region. This motif was additionally shown to be the binding motif for CreA/Cre-1 in *A. nidulans*, *T. reesei* and others (192, 193), however it is a poor predictor of actual CreA/Cre-1 binding sites (188, 190). This is likely due to its extremely short non-specific sequence. In *P. oxalicum,* CreA was shown to repress expression of *ClrB* and *XlnR* with experiments comparing expression of those two transcription factors using qRT-PCR in WT and a Δ*creA* background.

Here we explore the regulon of CRE-1 using DAPseq and show that CRE-1 mediated CCR acts not only through PCWDEs and their positive regulators, but through additional key alternative sugar catabolic genes, and importantly, a significant number of sugar transporters. If key sugar transporters are repressed, signal-transducing sugars cannot enter the cell, thus dampening expression of PCWDEs. Thus, CRE-1 regulates CCR through more than 4 levels of control or a

"quadruple lock" mechanism, which may explain the effectiveness of CRE-1 in preventing expression of PCWDEs.

5.5.2. *Results*

CRE-1 binds most frequently to promoters of carbon metabolism genes

To assess the direct binding sites of Cre-1, we conducted DAPseq using an N-terminal halo tagged, CRE-1 fusion protein with an indexed *N. crassa* FGSC 2489 genomic library. After sequencing, we filtered peaks called with Macs2 (181) in regions up to 3000BP upstream of the ATG start site of any gene. We found 329 peaks in 318 promoter regions, where 11 promoters regions had two peaks. We characterized the 318 genes with promoters bound by halo-CRE-1 using FunCat (121), and found 30 functional categories that were enriched with p-value < $1x10^{-5}$ (Table 5-4). The top 17 functional categories all involved carbon metabolism, specifically cellulose, hemicellulose, pectin and starch catabolism. Two further functional categories to note are the proton driven symporters and non-vesicular cellular import. The enrichment for these are likely due to binding of sugar transporters further discussed below.

Table 5-4 FunCat enrichment of CRE-1 binding peaks

Functional Category	p-value
01.05.03.06.07 hemicellulose/pectin catabolism	1.02E-19
01.05.03.06 hemicellulose/pectin metabolism	1.02E-19
01.05.03.06.07.02 specific pectin catabolism	9.98E-19
01.05.03.06.07.02.01 homogalacturonan catabolism	3.94E-18
01.05 C-compound and carbohydrate metabolism	5.67E-18
01.05.03.04.07 starch catabolism	1.82E-17
01.05.03.04 starch metabolism	1.82E-17
01.20 secondary metabolism	6.28E-17
01.05.03.06.07.03 hemicellulose/rhamnogalacturonan catabolism	1.20E-16
01.05.03.06.07.03.01 arabinan catabolism	3.10E-14
01.05.03 polysaccharide metabolism	5.93E-13
01.05.03.06.07.01.01 xylan catabolism	9.02E-13
01 METABOLISM	9.58E-13
01.05.03.06.07.01 specific hemicellulose catabolism	1.37E-11
01.05.03.05.07 cellulose catabolism	6.33E-11
01.05.03.05 cellulose metabolism	6.33E-11
01.06 lipid, fatty acid and isoprenoid metabolism	1.05E-09
20.09.18.07 non-vesicular cellular import	1.07E-09
20.03.02.02.01 proton driven symporter	5.50E-09
16.13.01 sugar binding	1.61E-08
01.05.02.07 sugar, glucoside, polyol and carboxylate catabolism	3.97E-08
20.03.02.02 symporter	9.56E-08

20.03.02 carrier (electrochemical potential-driven transport)	2.72E-07
34.01.01.03 homeostasis of protons	1.67E-06
42.01 cell wall	2.11E-06
20.01.13 lipid/fatty acid transport	2.30E-06
02.16 fermentation	2.92E-06
01.05.02 sugar, glucoside, polyol and carboxylate metabolism	3.06E-06
34.11.12 perception of nutrients and nutritional adaptation	3.62E-06
01.01 amino acid metabolism	5.89E-06
20.01.03.01 sugar transport	9.97E-06

CRE-1 binding is highly enriched for sugar transporters and to a lesser extent PCWDEs

We find that 18 MFS transporters are bound including all major cellodextrin transporters *cdt-1* (NCU00801), *cdt-2* (NCU08114) and *cdt-3* (NCU05853), arabinose transporter *lat-1* (NCU02188) and high affinity glucose transporter (NCU04963). The remaining transporters that are bound by CRE-1 according to our DAPseq data are shown in the table below. This is 18 out of 37 or 48% of the transporters identified in the MFS sugar clade. For reference, only 318/9700 or 3.2% of total genes are bound by CRE-1. Additionally, the glucose transporter (NCU04963) along with invertase (NCU04265) are the two genes that were shown to be bound by ChIP-PCR (190). Both of these genes are also bound within the same 200BPs within the promoter shown by ChIP-PCR, thus further affirming their authenticity as true CRE-1 binding sites. These data indicate that sugar transporters play a very important role in CRE-1 mediated CCR.

Table 5-5 MFS sugar transporter bound by CRE-1 DAP

Gene ID	annotation	Fold Enr	Dist. ATG
NCU04537	monosaccharide transporter	3.58	626
NCU01494	MFS sugar transporter	4.15	556
NCU00801	Cellodextrose transporter 1 *cdt-1*	4.64	326
NCU08114	Cellodextrose transporter 2 *cdt-2*	4.75	779
NCU00821	sugar transporter	6.38	2657
NCU06026	quinate permease	6.61	451
NCU06522	MFS maltose permease	7.55	648
NCU02188	L-arabinose transporter *lat-1*	8.74	574
NCU10021	MFS monosaccharide transporter	11.38	558
NCU01132	MFS monosaccharide transporter	12.95	788
NCU06384	MFS sugar transporter	13.75	504
NCU05627	high affinity glucose transporter *ght-1*	15.24	366
NCU09287	sugar transporter	15.32	897
NCU05853	Cellobionic acid transporter *cbt-1*	16.26	651
NCU04963	high-affinity glucose transporter	16.50	439

NCU00809	MFS monosaccharide transporter	17.82	339
NCU05897	l-fucose permease	38.96	970

For PCWDEs, we observe binding in the promoters of 19 out of ~120 enzymes in total (15%). These include the major exogalacturonanase (NCU06961), starch active LPMO (NCU08746), cellulose active PMO3 (NCU07760), and a number of xylanases, galactosidases and arabinases. We also observe several key sugar metabolic genes bound, including those crucial for arabinose and GalA utilization: LADH (NCU00643) and GaaA-C (NCU09533, NCU07064, NCU09532). The remaining PCWDEs that are bound by CRE-1 are shown in the table below (Table 5-6).

Table 5-6 PCWDEs bound by CRE-1 DAP

Gene ID	Annotation	Fold enr	Dist. ATG
NCU01900	intracellular beta-xylosidase	40.18	119
NCU08189	beta-1,4-endoxylanase	20.06	250
NCU07064	L-galactonate-dehdyratase GaaB	11.78	291
NCU09924	exoarabinanase	6.55	314
NCU09664	acetyl xylan esterase	6.47	356
NCU09923	extracellular beta-xylosidase	6.55	436
NCU08746	starch binding domain-containing protein	11.01	461
NCU06961	exogalacturonanase	14.09	463
NCU04623	extracellular beta-galactosidase	3.91	478
NCU08687	galactokinase	21.39	675
NCU04674	extracellular alpha-glucosidase	12.34	769
NCU00643	L-arabinitol-dehydrogenase LADH	9.18	810
NCU04265	Invertase	43.26	854
NCU09533	Galacturonic acid reductase GaaA	12.12	860
NCU06143	alpha-glucuronidase	7.28	929
NCU06523	intracellular alpha-glucosidase	7.55	1062
NCU01517	extracellular alpha-glucosidase	7.38	1063
NCU00642	extracellular beta-galactosidase	9.18	1299
NCU00943	trehalase	7.71	1405
NCU09532	L-threo-3-deoxy-hexulosonate adolase GaaC	12.12	1510
NCU07760	polysaccharide monooxygenase 3 (PMO3)	4.43	2428

Surprisingly, we only detected binding of three described transcription factors bound by CRE-1. These were *clr-1* (NCU07705) and *nit-4* (NCU08294) and *ara-1 (*NCU05414). *Nit-4* is interesting in that if the interaction between CRE-1 and *nit-4* is authentic, it indicates a connection between carbon catabolite repression and nitrogen metabolism. Additional genes involved in phenylalanine catabolism and branched chain amino acid metabolism are also bound by CRE-1. These include fumarylacetoacetase (NCU05537), homogentisate 1,2 dioxygenase (NCU05499),

branched chain alpha-keto acid dehydrogenase E2 component (NCU02704) and methylcrotonoyl-CoA subunit alpha (NCU00591). These are all bound within 1kb of their ATG start sites. Coincidentally, CreA in *A. nidulans* also plays a role in amino acid transport and nitrogen metabolism (194).

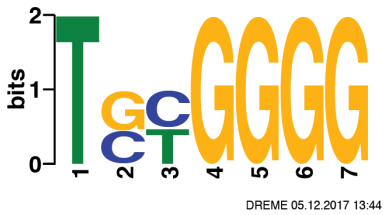
All in all carbon metabolism genes represent about a half of the total peaks detected. Whether non-carbon metabolism binding sites we observe are authentic and or functional in an *in vivo* setting is unknown. We do not have any hypothesis regarding whether these binding sites are important for CRE-1 function.

CRE-1 binding motif built from DAPseq peaks agree with previously described motifs.

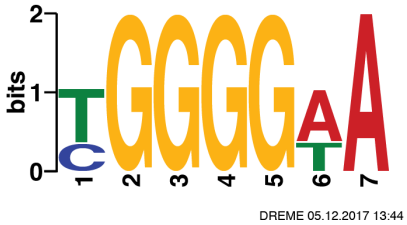
We used the sequences from 232 CRE-1 binding peaks detected in promoter regions to build a consensus core motif using DREME v4.12.0 (184). The core motif with the lowest E-value ($2.7x10^{-23}$) was 5'-TSYGGGG-3' (Figure 5-12A). This is very similar to the CreA motif described in *Aspergillus nidulans,* 5'-SYGGRG-3' (193). However, a second motif with low E-value ($2x10^{-11}$) was also found. This motif shares four consecutive guanines with the first motif, but has two additional bases after guanines biased for adenine (Figure 5-12A). We also built a third motif using only the 17 peaks found within sugar transporter genes using MEME v4.12.0 (184). This yielded a combination of the two core motifs, with three additional guanines several bases toward the 3' end (Figure 5-12B).

A

CRE-1 core motifs constructed with DREME (All 232 unique peaks)



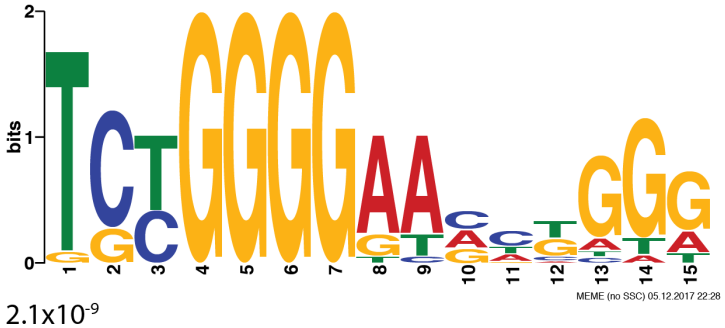
E-value $2.7x10^{-23}$



E-value $2.0x10^{-11}$

B

CRE-1 MFS sugar transport motif constructed with MEME (from 17 unique peaks)



E-value $2.1x10^{-9}$

Figure 5-12 CRE-1 binding motifs. *A) top two motifs constructed from 232 unique DAPseq peak sequences found in promoter regions (3kb from any ATG start site) using DREME v4.12.0. B) Consensus motif constructed using only peak sequences from sugar transporters bound by CRE-1. Motif built with MEME v4.12.0.*

5.5.3. ***Discussion***

The results of our direct CRE-1 binding data can be interpreted in several ways. The first assumes that we have correctly identified the majority of true binding sites for CRE-1. The second assumes that we are missing a large number of true binding sites, and the sites we detected are only a portion of the total CRE-1 bound sites. Arguments can be made for or against these assumptions and this discussion focuses on the conclusions that can be drawn if either of these assumptions are correct. A third scenario is one in which many of our called peaks are actually false positives. We believe this to be unlikely due to enrichment of carbon metabolic genes, which we know CRE-1 must act through.

If we assume that we have found the majority of CRE-1 binding sites, then we can say that CRE-1 repression of PCWDE occurs through a mechanism is not in accordance with previously published CRE-1 data. Our data is unanticipated in that the only carbon TF that has a CRE-1 binding peak is *clr-1*. This would mean that the rest of the carbon catabolism TFs notably, *xlr-1, clr-2, pdr-1, pdr-2, ara-1* are not directly repressed by Cre-1. Additionally very few extracellular PCWDEs are

directly bound by CRE-1, only 19 out of approximately 110 in total. Major cellulases, LPMOS and xylanases showed no CRE-1 binding, including CBH-1 and others with CRE-1 binding evidence described in previous literature (190). Nor is the complete set of alternative sugar catabolism genes bound by CRE-1. Only arabinose, GalA and galactose catabolism would be disabled under CRE-1 mediated CCR. No other sugar catabolic gene promoters are bound. Neither xylose nor rhamnose catabolic genes, nor any β-glucosidases were bound. However, 18 of the 37 sugar transporters in the *N. crassa* genome are bound. All of the known sugar transporters fall into a subclade within the MFS superfamily, and this subclade consists of only 37 transporters (Figure 3-15). Furthermore, 6 of these 37 transporters are high affinity glucose transporters that are not bound by CRE-1. This leaves us with 16 sugar transporters. A few of these can be eliminated as transporters of signal inducing sugars such as quinate permease (NCU06138) ~14 sugar transporters that may or may not be transporting significant amounts of signal inducing sugars. It is therefore a possibility that CRE-1 acts mainly through reducing the signal to induce expression of PCWDE by reducing the amount of inducing sugars that enter the cell. This seems especially viable with cellulase production, as all three CDTs are bound by CRE-1, thus providing weak or no signal in the cell to activate CLR-1 and its downstream cellulases.

If we assume that we are indeed missing a good number of binding sites of CRE-1, transporters may not be the main avenue of PCWDE repression by CRE-1. It has already been shown that CRE-1 binds to the promoters of both *xlr-1* and xylanases in other species of fungi (187, 189), and it is very possible that our DAPseq did not pick these peaks up. DAPseq is, after all, an in vitro technique; CRE-1 could be binding these genes but not observed in DAPseq results for a number of reasons. It was also shown in Neurospora that 9 genes showed enrichment by ChIP-PCR using CRE-1 protein as bait (190). Only 2 of these genes overlap with our DAP results, indicating one or both of the techniques failed. It is also possible that DAPseq only picked up the strongest binding sites in the genome. After all, the set of CRE-1 bound genes by DAP are relatively small, only 318 promoters in total. One additional set of experiments indicates that CRE-1 does repress expression of PCWDEs to a greater extent than as indicated by DAPseq. When *clr-2* is overexpressed under sucrose conditions, cellulase production is still repressed as measured by cellulase activity in the supernatant (28). RNAseq also shows that transcription of cellulases is repressed in the *clr-2* mis-expression strain under sucrose conditions (data not shown). Whereas the *clr-2* mis-expression strain under no carbon conditions produces as much cellulase as WT on Avicel. These data indicate that there is some mechanism repressing cellulase transcription under sucrose that can override mis-expression of *clr-2.* The likely repressor is CRE-1.

Regardless of which assumption is correct, our data indicates that the non-glucose sugar transporters in the *N. crassa* genome certainly have a high likelihood of being bound by CRE-1 as observed through DAPseq. These transporters likely are involved in dampening the signal to positive regulators that are responsible for increasing expression of the major PCWDEs in *N. crassa.* Whether or not they are the main contributor to CRE-1 mediated CCR requires further investigation.

CRE-1 connection to nitrogen metabolism needs to be fully explored. We show that CRE-1 is bound to *nit-4* promoter, and a few enzymes involved in branched chain amino acid metabolism and phenylalanine metabolism. NIT-4 is responsible for co-regulating expression of nitrate reductase and another aspects of nitrogen metabolism. Evidence from *A. nidulans* support that authenticity of this binding peak (194).

5.6. Expanding the regulon of NIT-2 and its role in nitrogen metabolism

5.6.1. *Introduction*

NIT-2 is one of the main transcription factors involved in nitrogen metabolism in *N. crassa*. With the addition of NIT-4 and NMR, *N. crassa* is able to distinguish and metabolize different nitrogen sources depending on the presence and absence of others. The interaction of these three transcription factors has been studied by using the system of *nit-3* (nitrate reductase) expression as a model for other genes. *N. crassa* prefers ammonia, glutamate and glutamine over nitrate. In the absence of those three nitrogen sources, *N. crassa* is able to efficiently utilize nitrate to fulfill its nitrogen requirements. NIT-2 and NIT-4 work cooperatively to induce expression of *nit-3*, which is a required enzyme for nitrate utilization. Nitrate is then converted to nitrite and in turn ammonia, which is further metabolized to generate glutamate and glutamine. NMR (nitrogen metabolite regulator) is a repressor of NIT-2, and is active when ammonia/glutamine/glutamate are in high abundance, thus preventing expression of *nit-3* (see section 3.6.3 for more information). Less information is known outside this model gene system, especially regarding the full regulon of NIT-2 and NIT-4 (150). We were able to use DAPseq to explore the full regulon of NIT-2. We find that our data is in accordance with previous literature, and that our RNAseq and DAPseq data support each other.

5.6.2. *Results*

We conducted DAPseq on NIT-2 and NIT-4 to determine the role of each transcription factor in regard to nitrogen metabolism. Although much research has been conducted to show how both interact and induce expression of a few enzymes, the general regulon of these two TFs has not been fully explored. Unfortunately, we did not get any significant bindings peaks from NIT-4 DAPseq, indicating that either our experiment failed on a technical basis, or that NIT-4 requires other partners for DNA binding. We were, however, able to detect a number of binding sites of Nit-2 through DAPseq. We found that our peaks generally reflected RNAseq results comparing WT to *Δnit-2*, thus indicating that a portion of the peaks we observed are true positives.

Overview of RNAseq results are in accordance with previous literature regarding NIT-2 involvement in nitrogen metabolism

We found in our RNAseq data that the set of genes is down regulated in the Δ*nit-2* compared to WT on Vogel's minimal media (ammonium nitrate) that are related to acquisition of nitrogen. The top gene down regulated is agmatinase, which cleaves agmatine into urea. Many nitrogen acquisition genes are down regulated including urea permease, nitrate transporter, ammonium transporter, peptide transporters, amino acid transporters and serine peptidase. Of these genes, only the nitrate transporter (NCU07205) has been shown to be regulated by NIT-2 (195). *Nit-3* and *Nit-6 (*nitrite reductase) anabolic genes are also down regulated, which have previously been shown to be regulated by NIT-2 (51, 196).

NIT-2 DAPseq results are in accordance with previously described NIT-2 regulon and coincide with RNAseq data

Overall we found 344 genes bound by halo-NIT-2. Ten of these genes had a secondary binding site in their promoter region. About half of these (154 genes) are hypothetical proteins. We speculate that a number of these genes are false positives, as often observed with ChIPseq results (31). To determine which genes are true positives, we overlayed our DAPseq data over our RNAseq differential expression data between WT and Δ*nit-2* mutant on Vogel's media containing ammonium nitrate as the nitrogen source with 2% sucrose. We found that about half of the most highly down regulated genes in Δ*nit-2* are also bound by NIT-2 (Figure 5-13). If we examine the top differentially genes more closely, we noticed that a number of these genes are not bound (Table 5-7). This indicates one of the three possible interpretations: 1) Promoters of these genes are not directly regulated by NIT-2, but rather another TF downstream of NIT-2. 2) NIT-2 requires additional partners in order to bind these genes. 3) Binding is occurring *in vivo* but not detected by DAPseq. Further discussion of possible interpretations are discussed below.

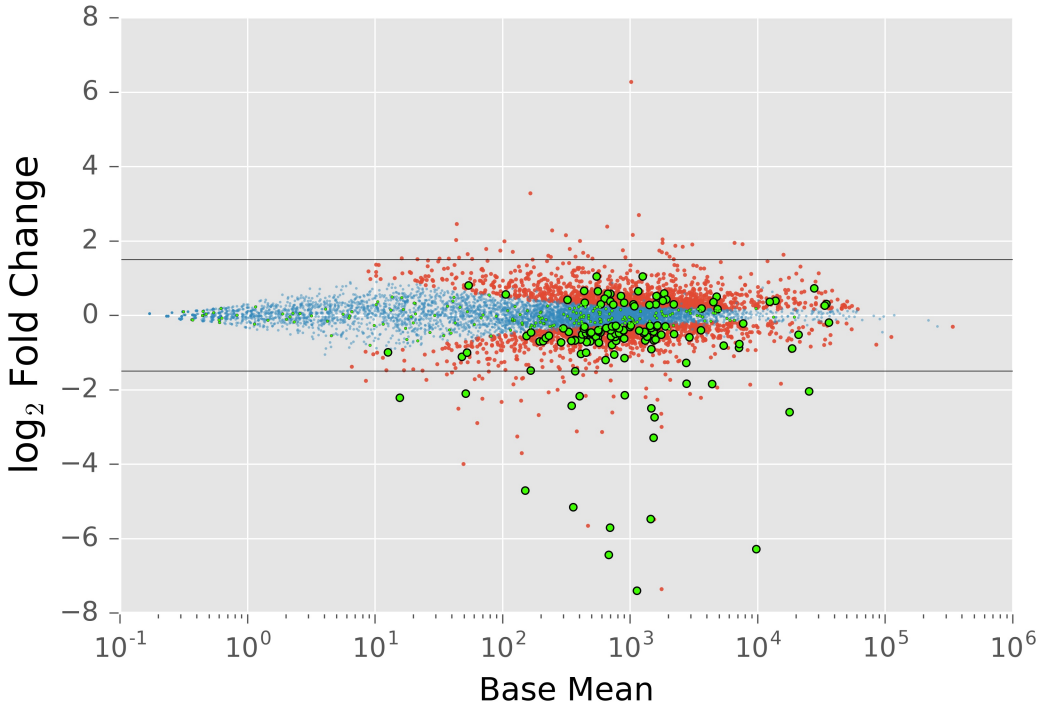


Figure 5-13 Differential expression plot of WT vs Δ*nit-2* on Vogel's (ammonium nitrate) with 2% sucrose, NIT-2 DAP peaks highlighted. *Genes highlighted in blue represent genes differentially expressed with adjusted Pvalue >.01, and red <.01. Genes in green have promoter region up to 3KB bound by Nit-2, those outlined in black are significantly differentially expressed (adj. Pval <.01).*

Table 5-7 Top differentially down regulated genes in expression profiles of the *Δnit-2* strain combined with DAPseq binding data

Gene ID	baseMean	log2FoldChange	Nit-2 bound	Annotation
NCU01348	1128.22	-7.39	YES	agmatinase
NCU05298	1756.60	-7.36	no	nitrate nonutilizer-3, nit-3
NCU06416	678.64	-6.43	YES	thymine dioxygenase
NCU07205	9713.42	-6.28	YES	nitrate transporter CRNA
NCU08407	693.20	-5.70	YES	MFS transporter
NCU07894	466.96	-5.65	no	oligopeptide transporter 2
NCU08397	1541.69	-5.48	no	hypothetical protein
NCU01065	1442.07	-5.47	YES	ammonium transporter MEP2
NCU09909	356.24	-5.15	YES	urea active transporter
NCU04435	150.37	-4.71	YES	general amino acid permease AGP3

NCU02061	49.15	-3.99	no	hypothetical protein
NCU07923	140.93	-3.70	no	hypothetical protein
NCU08225	1519.12	-3.29	YES	high affinity nickel transporter nic1
NCU11405	130.26	-3.25	no	α-1,3-glucanase/mutanase
NCU09992	599.80	-3.13	no	serine peptidase
NCU04430	380.80	-3.11	no	leupeptin-inactivating enzyme 1
NCU09788	62.96	-2.89	no	endonuclease/exonuclease /phosphatase
NCU07334	1545.12	-2.74	YES	uracil permease
NCU09506	191.13	-2.67	no	hypothetical protein
NCU10276	1749.01	-2.64	no	hypothetical protein
NCU05079	721.53	-2.61	no	MFS peptide transporter
NCU04720	17738.92	-2.60	YES	nitrate nonutilizer-6, nit-6

NIT-2 binding motif

We used the binding peaks from this list of NIT-2 direct targets to identify a binding motif. By taking the sequences of the binding peaks, we used MEME v4.12.0 (184) to build a consensus motif common within all input sequences. A single motif with a significant E-value of 3.3×10^{-6} was discovered (Figure 5-14).

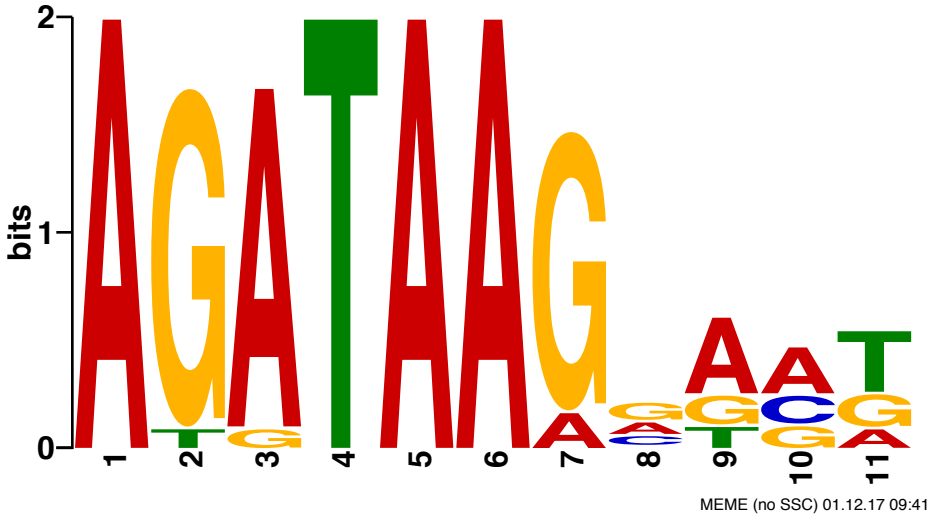


Figure 5-14 NIT-2 DNA binding motif

NIT-2 may require binding partner NIT-4 for binding of some genes as hypothesized in previous literature

It is likely that NIT-2 requires additional partners for binding. A previous study examined promoter mechanism for *nit-3 (*nitrate reductase) that is highly down regulated in *Δnit-2* strain and a known target of NIT-2 positive regulation (149, 150). This study showed that *nit-3* promoter region has two distal binding

sites adjacent to one another about 1kb upstream of the ATG start site (151). These are directly upstream of two *nit-4* binding sites. Both TFs are required for expression and production of NIT-3 (150, 151). Furthermore, Chiang and Marzluf et al. 1995 (151) showed that when any of these four binding sites were mutated, the production of NIT-3 was greatly reduced. The two binding sites in the middle, one for NIT-2 and one for NIT-4, are especially important. When either of those are mutated, there is no detectable production of NIT-3. Authors Chiang and Marzluf (151) hypothesize that two transcription factors bind cooperatively, and that binding of one increases affinity of binding for the other (151). If true, this could explain why our DAPseq of NIT-2 did not show any binding of the NIT-3 promoter region. Research regarding promoter regions of other NIT-2 targets has not been conducted. In summary, we hypothesize that some of the genes down regulated in the *Δnit-2* strain do not show NIT-2 binding because NIT-4 is also required for binding those sites.

One caveat that remains, however, is that this same research discovered a third NIT-2 binding site about 300BP upstream of ATG that was shown that it can be bound to NIT-2 through EMSA. These binding conditions are different from our DAPseq conditions, and may be sensitive enough to pick up low affinity binding not detected in our data.

5.6.3. ***Summary and Discussion***

We found that the NIT-2 regulon as determined by RNAseq was supported by DAPseq data. Genes involving nitrogen acquisition that we expected to be down regulated in the Δ*nit-2* strain were indeed down regulated according to our RNAseq data. Many of these genes were also bound by NIT-2 indicating that they are directly regulated by NIT-2 rather than a downstream mechanism. However, many genes did not have corresponding binding data. We reason that these genes may require NIT-4 for binding, as hypothesized previously with promoter binding of NIT-3.

Many caveats remain regarding full NIT-2 regulon especially with the interaction of NIT-4. These two transcription factors have already been shown to interact cooperatively. It would be interesting to compare our data with data from DAPseq of NIT-2 and NIT-4 proteins together. DAPseq has especially good potential for study interacting transcription factors, as protein from two TnT experiments can easily be combined before addition of DNA library. Furthermore, additional RNAseq experiments could be done to maybe find full NIT-2 and NIT-4 regulons. The conditions we used for our Δ*Nit-2* RNAseq used Vogel's media, which contains both ammonia and nitrate, and in retrospect we likely should have used conditions where nitrate was the sole nitrogen source.

5.7. Exploring direct binding partners of NUC-1 and NCU03077 using DAPseq

5.7.1. ***Introduction***

We used DAPseq to further explore the direct binding partners of NUC-1 and NCU03077. The regulon of both of these transcription factors is explored in section 4.6, and experiments presented in this section follow the data presented there. DAPseq results from NUC-1 confirm several direct targets discovered in the past, and also provide a novel binding motif for NUC-1. DAPseq results from NCU03077 are uninformative regarding the role of this transcription factor.

5.7.2. ***Results***

NUC-1 DAPseq confirms small number of NUC-1 targets, but many expected targets are missing

NUC-1 DAPseq resulted in 282 unique peaks within 3KB promoter regions of 381 genes. When we overlayed the DAPseq data onto RNAseq differential expression analysis of WT vs *Δnuc-1* we were able to confirm that a number of the differentially expressed genes are direct targets of NUC-1 (Figure 5-15A). We conclude that these genes greater than 4 fold downregulated in the *Δnuc*-1 strain and also bound by NUC-1 DAPseq are true positives and direct targets of NUC-1. We counted false positives as genes bound by NUC-1 but show no differential expression in the RNAseq. False positive rates were no higher than other DAPseq data sets. However, we noted a number genes we believe to be false negatives. These genes are highly downregulated in *Δnuc-1* but do not appear to be bound by NUC-1 according to DAPseq. These are displayed in Table 5-8, which shows the top downregulated genes in Δ*nuc-1*. The unbound genes in this list involved in phosphate acquisition are our top false negative candidates. When we plot the percentage of genes bound with increasing $\log_2$ fold change threshold, we observe a positive slope for genes downregulated in *Δnuc-1* strain as expected from a positive regulator like NUC-1. However the percentage of bound genes never breaches 50%. This is further indication of the abundance of false negatives (Figure 5-15B).

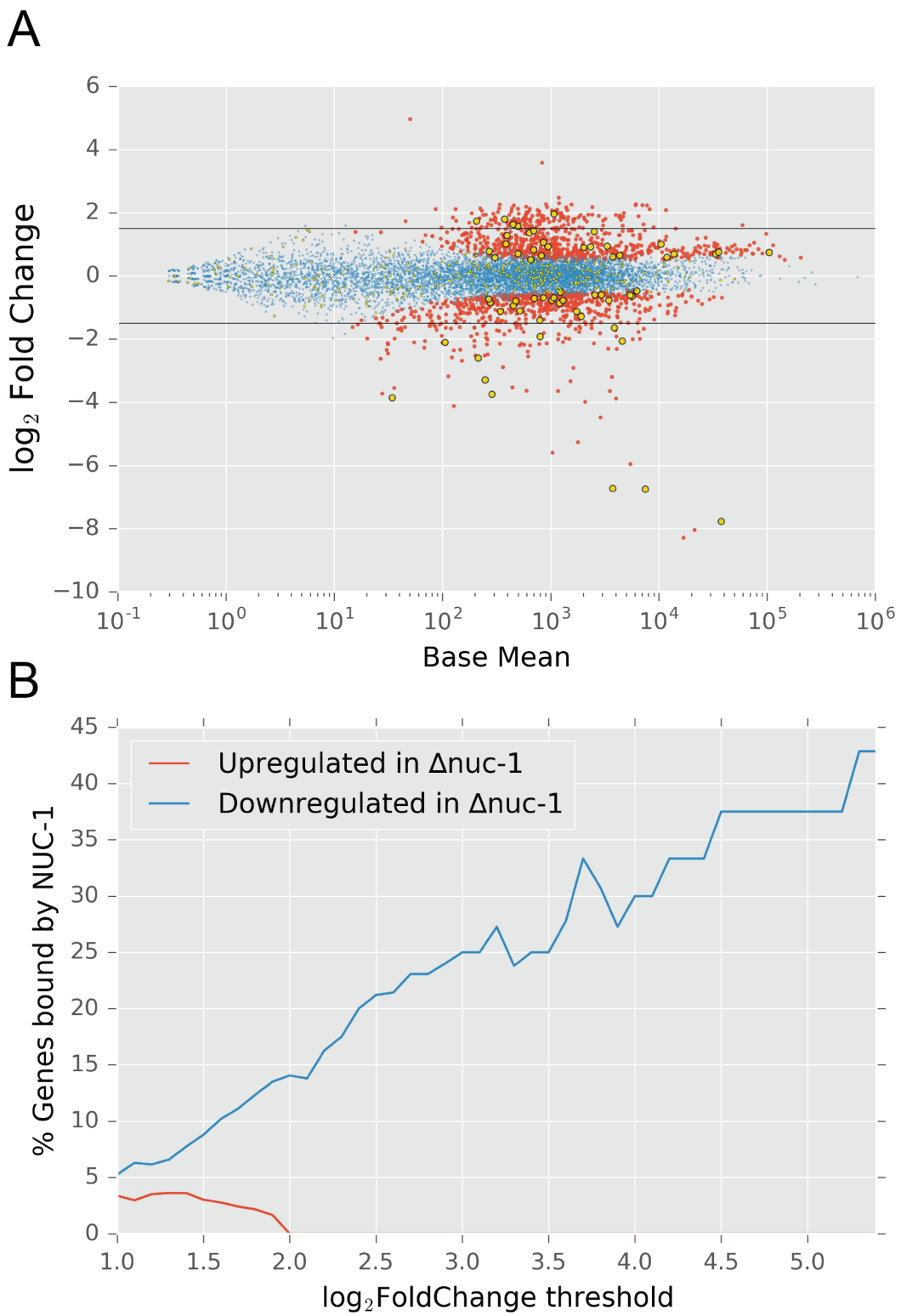


Figure 5-15 NUC-1 DAPseq overlayed onto differential expression of WT vs *Δnuc-1* no phosphate. *A) genes highlighted in blue represent genes differentially expressed with adjusted P value >.01, and red <.01. Genes in yellow have promoter region up to 3KB bound by Nit-2, those outlined in black are significantly differentially*

expressed (adj. Pval <.01). B) plot of % genes bound as a function of log_2 fold change threshold for upregulated and down regulated genes.

Table 5-8 Most downregulated genes in *Δnuc-1* compared to WT on no phosphate with binding from DAPseq.

Gene ID	baseMean	log2FoldChange	fold enr	dist ATG	annotation
NCU08325	16878.72	-8.29	not bound	not bound	phosphorus-5, pho-5
NCU09564	21390.19	-8.04	not bound	not bound	phosphorus-4, pho-4
NCU01376	37787.82	-7.77	5.78	326	phosphorus-2, pho-2
NCU10038	7501.32	-6.75	9.12	255	glycerophosphoryl diester phosphodiesterase
NCU08643	3756.27	-6.73	7.79	2763	acid phosphatase
NCU03032	5454.49	-5.95	not bound	not bound	betaine lipid synthase
NCU09631	1038.35	-5.59	not bound	not bound	hypothetical protein
NCU09767	1786.05	-5.26	not bound	not bound	membrane transporter
NCU09969	66.98	-5.08	not bound	not bound	hypothetical protein
NCU09525	2871.72	-4.48	not bound	not bound	secreted protein
NCU09630	127.38	-4.11	not bound	not bound	hypothetical protein
NCU03077	2083.53	-3.99	not bound	not bound	hypothetical protein
NCU04809	4013.77	-3.88	not bound	not bound	MFS phospholipid transporter
NCU07486	34.39	-3.86	9.58	1178	LysM domain-containing protein
NCU07485	285.21	-3.74	4.27	1703	hypothetical protein
NCU03813	3514.49	-3.64	not bound	not bound	formate dehydrogenase
NCU01045	1169.96	-3.64	not bound	not bound	ribonuclease T1
NCU09315	597.11	-3.63	not bound	not bound	nuclease-l, nuc-1
NCU07449	35.63	-3.55	not bound	not bound	hypothetical protein
NCU04533	442.77	-3.53	not bound	not bound	abundant perithecial protein, app
NCU01055	1520.03	-3.33	not bound	not bound	chromate ion transporter
NCU07133	247.18	-3.29	4.50	1070	metallo-beta-lactamase superfamily protein
NCU03415	3658.89	-3.19	not bound	not bound	aldehyde dehydrogenase
NCU09761	112.92	-3.18	not bound	not bound	hypothetical protein

NUC-1 consensus binding motif

With the low number of direct targets genes (9 genes), we were unable to construct a significant binding motif. Instead we reduced the stringency of differential expression fold change to >2x in order to build a statistically significant motif. This new set of peaks fell within the promoters of 16 genes. Nucleotide sequences of these peaks were used to construct a single statistically significant DNA binding motif with the E value of 8.1×10^{-4} (Figure 5-16) using MEME v4.12.0 (184).

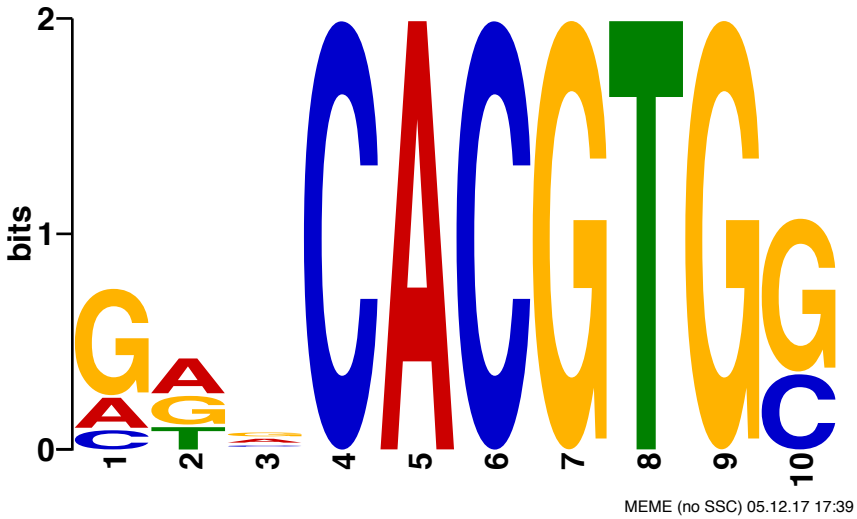


Figure 5-16 NUC-1 binding motif

Furthermore, this motif was found in the center every input peak sequence. Such an observation should only occur if the motif is authentic. This is because reads fall around the true binding site of the transcription factor thus the center of the peak is the only location where all reads overlap. If the motif is also located in this region, than it is more likely that the motif is authentic. We made this observation with the motif that we found (Figure 5-17).

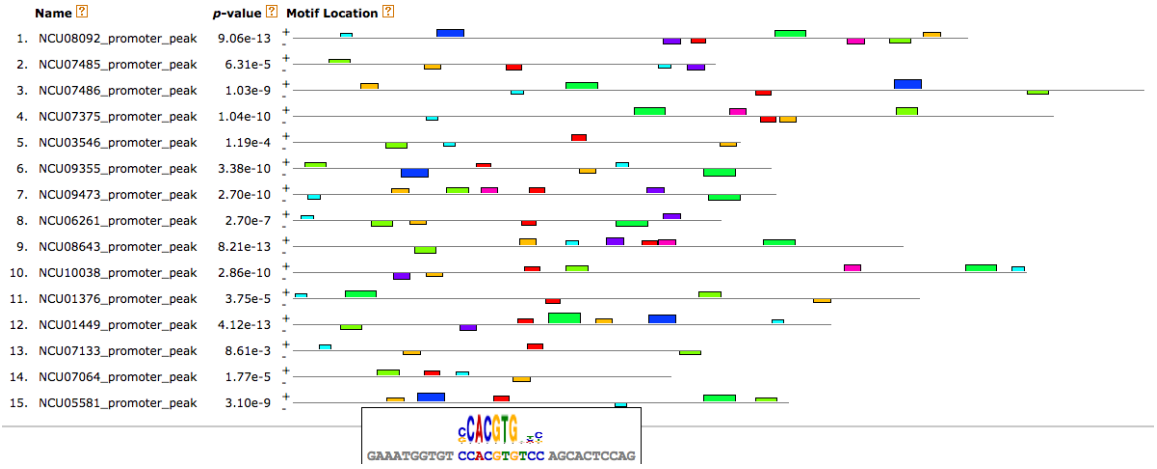


Figure 5-17 Motif location within DAP peaks. *Screenshot of MEME output where the name of each input promoter peak is labeled on the left. The position of CACGTG motif within each peak sequence is shown in red. Note that the motif falls close to the center of each peak sequence and is also present in all sequences.*

NCU03077 DAPseq reveals binding motif but uninformative for TF function

DAPseq of NCU03077 yielded 1392 unique peaks. These peaks fell within the within 3kb of the ATG start site of 1794 genes. This is approximately 20% of the genes in the *N. crassa* genome. We are unable to identify any biologically relevant

peaks with confidence, as RNAseq on NCU03077 yielded very few differentially expressed genes (see section 4.6.2).

We were, however, able to build a robust binding motif using sequences from all binding peaks that fell within promoter regions. We used DREME v4.12.0 to construct a core DNA binding motif. A single motif with an E-value of 1.8×10^{-197} is likely the true NCU03077 binding site (Figure 5-18), with the next most significant motif having an E value of 1.2×10^{-33}.

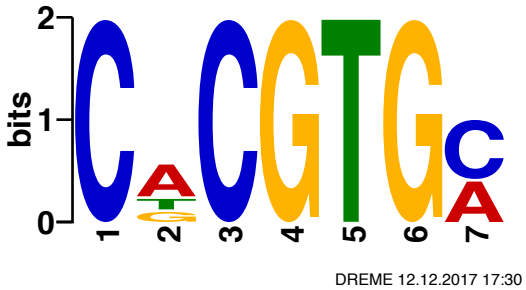


Figure 5-18 NCU03077 DNA binding motif. *This motif was built using DREME v4.12.0 with maximum core width set to 10 base pairs. All other parameters set to default.*

5.7.3. *Conclusions*

We conclude that the DAPseq of NUC-1 is missing many genes. This may be due to in vivo binding partners absent from the DAP conditions, or other unknown aspects of the experiment. However, we were still able to construct a robust binding motif. Future analyses will focus on deterring presence or absence of this motif in candidate NUC-1 target genes missed by DAPseq.

NCU03077 yielded many binding sites, but none can be verified by RNAseq. Further exploration of the conditions in which NCU03077 might be of more importance are required to gain a better understanding of the role of this transcription factor. A robust binding motif built from DAPseq peaks may be of use to future researchers investigating this TF.

5.8. Expanding the CYS-3 regulon and its role in sulfur starvation response

5.8.1. *Introduction*

As discussed in section 3.6.2, CYS-3 is only known TF involved in response to sulfur limitation. Although the full mechanism is unknown, CYS-3 is activated under sulfur limiting conditions. Once active, CYS-3 induces production of key sulfur acquisition structural genes including a number of sulfatases and sulfur permeases (147). We conducted DAPseq to assay the direct binding sites of CYS-3 within the genome, in order to try and better understand CYS-3 role in sulfur starvation response.

5.8.2. ***Results***

Confirmation and expansion of CYS-3 regulation of sulfur acquisition genes.

From our DAPseq analysis, we found 669 genes with promoter regions (up to 3KB upstream) bound by halo-CYS-3. If we examine only the known sulfur acquisition genes in the genome as annotated in Borkovich et al. 2004 (147), we see that the vast majority are bound by CYS-3. We show that CYS-3 does indeed bind to the promoters of many of these genes. Some of these genes have yet been shown to be directly regulated by CYS-3 such as the taurine dioxygenases and methionine permeases. We find that the majority of these genes are also directly bound by halo-CYS-3. This information is summarized in Table 5-9, which includes peak enrichment and the distance of the peak maxima from the ATG start site of each gene. Promoters not bound by halo-CYS-3 display "N/A" in these two columns. Of the known sulfur acquisition genes, only one taurine dioxygenase, cysteine dioxygenase, sulfite oxidase, and two sulfate permeases are not bound by CYS-3. These genes have never been shown to be under regulation by CYS-3.

Table 5-9 Sulfur acquisition genes bound by halo-CYS-3. *All sulfur acquisition genes summarized in Borkovich et al. 2004* (147) *listed in this table. Genes highlighted in green represent sulfur acquisition genes, red represent transporters and orange represent gene regulators. Fold enrichment is calculated by MACS2 peak calling software.*

Gene ID	Broad v12 Annotation	Distance to ATG (BP)	fold enr.
NCU06041	Arylsufatase (ARS-1)	377	40.03
NCU08364	Choline sulfatase	350	16.73
NCU10015	Alkanesulfonate monooxygenase	311	46.45
NCU05340	Alkanesulfonate monooxygenase	346	38.35
NCU06112	Cysteic acid decarboxylase	528	7.85
NCU09738	Taurine dioxygenase	N/A	N/A
NCU09800	Taurine dioxygenase	378	11.57
NCU07610	Taurine dioxygenase	550	19.13
NCU07819	Taurine dioxygenase	770	8.70
NCU06625	Cysteine dioxygenase	N/A	N/A
NCU04474	Sulfite oxidase	N/A	N/A
NCU06931	Sulfite oxidase	N/A	N/A
NCU03235	Sulfate permease I (CYS-13)	1443	34.97
NCU04433	sulfate permease II (CYS-14)	1334	12.99
NCU09642	Sulfate permease	N/A	N/A
NCU02632	Sulfate permease	N/A	N/A
NCU02195	Methionine permease	2503	31.65
NCU04942	Methionine Permease	709	20.22
NCU07754	Methionine permease	314	15.08
NCU01055	Chromate resistance efflux	352	20.39

NCU08563	Scon-2	1561	33.99
NCU03536	Cys-3	N/A	N/A

CYS-3 also binds promoters of genes downstream of sulfur acquisition

We found a number of additional sulfur metabolic genes to be bound by CYS-3, not directly related to sulfur acquisition. These genes are involved in general cysteine and sulfide metabolism, as well as homocysteine and methionine metabolism. These are summarized in Table 5-10. In order to for sulfate to be assimilated in to cysteine, the cell converts sulfate to APS (adenosine-5'-phosphosulfate) to PAPS (3'-phosphoadenosine-5'-phoshosulfate) to sulfite to sulfide to cysteine. ATP sulfurylase (CYS-11) converts sulfate to APS, and PAPS reductase (CYS-5) converts PAPS to sulfite. Below is a table of the sulfur metabolism genes bound by CYS-3. Genes in blue are involved in the inter-conversion of homocysteine and methionine (147). This data indicates that CYS-3 is involved varying aspects of sulfur metabolism in addition to sulfur acquisition.

Table 5-10 CYS-3 bound cysteine, sulfide, homocysteine and methionine metabolism genes. *Genes highlighted in purple represent sulfide and cysteine metabolism, while blue represents homocystein and methionine metabolism. Fold enrichment and location of peak maxima calculated by MACS2 2 software.*

Gene ID	Broad v12 Annotation	Distance to ATG (BP)	fold enr.
NCU01985	ATP sulfurylase (CYS 11)	1246	4.12
NCU02005	PAPS reductase (CYS 5)	517	6.40
NCU01652	O-acetyl-homoserine sulfhydrolase	798	5.73
NCU09545	Methylene tetrahydrofolate reductase	1243	24.40
NCU08216	Cystathionine β synthase	490	12.11
NCU09230	Cystathionine gamma lyase	383	43.58

MFS transporter candidates found up regulated on sulfur also bound by CYS-3

MFS transporter candidates that found to be up regulated on no sulfur are also shown to be bound by CYS-3. We found 6 MFS transporters that were up regulated solely under no sulfur conditions (see section 3.5.2). None of the carbon conditions we tested displayed any expression of these genes, and only under sulfur starvation were these genes expressed to any reasonable degree (Figure 5-19). Several of these genes have annotations derived from homology to other proteins including NCU05884 and NCU07820, which are SEO1 (Suppressor of sulfoxyde EthiOnine resistance) and pantothenate transporter, respectively. Two of the other candidates contain DUF895 domains (NCU05886, NCU10512). All 6 of these genes are bound by CYS-3 according to our data (Table 5-11). Interestingly, transporter NCU06040 is located next to aryl-sulfatase, a key structural gene for sulfur acquisition discussed above. The two genes share a promoter region, but are transcribed in opposing directions. It is possible that the two represent a gene cluster important for sulfur scavenging. It is also possible that NCU06040 induction is a byproduct of aryl-sulfatase promotion by CYS-3 that is biologically irrelevant.

The two transporters SEO-1 (NCU05884), and un-annotated transporter NCU05886 also share a promoter region.

Table 5-11 binding peaks of MFS no sulfur response candidates

Gene ID	Broad v12 Annotation	Distance to ATG (BP)	fold enr.
NCU06040	hypothetical protein	442	40.03
NCU07820	pantothenate transporter	388	8.70
NCU09039	hypothetical protein	1650	32.82
NCU10512	DUF895 domain membrane protein	740	7.78
NCU05884	MFS transporter Seo1	998	32.75
NCU05886	DUF895 domain membrane protein	491	6.64

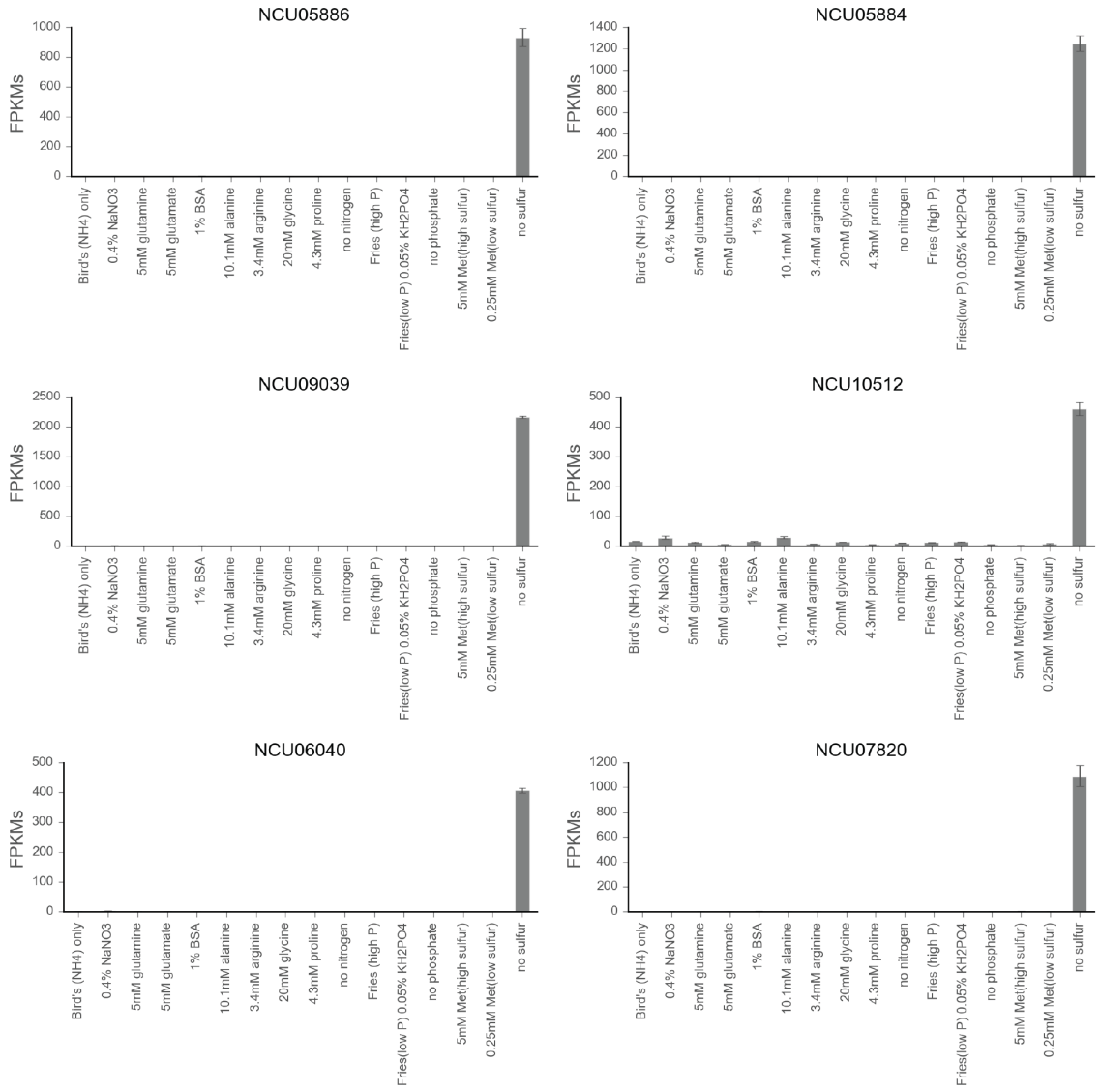


Figure 5-19 Expression profile of MFS no sulfur response candidates.
Normalized counts are plotted for the expression of each gene across our NPS

conditions. Carbon conditions not included due to lack of expression in those conditions. Error bars represent standard deviation of biological triplicates.

Role of CYS-3 in Auto regulation

Contrary to previous literature, we did not find that CYS-3 is auto regulatory. A study by Paietta, 1992 demonstrated that a CYS-3 promoter fused to LacZ gene was induced when *cys-3* was overexpressed, thus indicating that CYS-3 could induce expression of itself (197). We did not find any binding peaks for halo-CYS-3 in the *cys-3* promoter.

We did, however find that *scon-2* (NCU08563) was bound by CYS-3, which is in accordance to the model for interaction of CYS-3 by the SCON-2 complex (145). Although, SCON-2 is a repressor for CYS-3, CYS-3 is required for expression of SCON-2 in a negative feed back loop. This likely ensures enough SCON-2 is present for full repression of CYS-3 when needed. A similar system of repression is seen with transcription factors NUC-1, PDR-2 and the cooperative CLR-1 and CLR-2 (39, 183).

CYS-3 binds to the promoters of a large portion of genes up regulated under sulfur starvation

In order to further understand the role of CYS-3 in sulfur starvation response, we overlayed our CYS-3 DAPseq data onto our differential expression analysis of wild type cells between high sulfur and no sulfur conditions. As we discussed in section 3.6.2, there is an enormous number of genes differentially expressed between those two conditions, both up and down regulated. We plotted our differential expression data highlighted with the genes bound by CYS-3 according to our DAPseq data. There are 2678 genes significantly up regulated and 2736 genes significantly down-regulated with adjusted p value less than .01. Of the up-regulated genes, 11.38% of the genes are also bound by CYS-3. Of the down-regulated genes 5.11% are bound (Figure 5-20A). We noticed that if we increase the threshold of differential expression significance to adjusted P value $>1x10^{-9}$, the percentage of genes bound increases for genes that have increased expression on sulfur starvation. The percentage rose from 11.38% to15.18%. Contrastingly, the percentage of genes that displayed reduced expression under sulfur starvation remained at ~ 5% (4.79%) when p value threshold was as increased to $1x10^{-9}$.

We then plotted the percentage of genes bound as a function of significance to determine how this number bound by CYS3 changes with increasing significance. We found that genes that are more significantly differentially regulated are more likely to be bound by CYS-3 but only if they have increased expression on sulfur limitation. If their expression decreases on sulfur starvation, they have the same likelihood of being bound by CYS-3 regardless of how significantly they are down regulated (Figure 5-20C). The same pattern is observed if we use $\log_2$ fold change as the threshold instead of adjusted P value (Figure 5-20D). The fraction of genes bound by CYS-3 remains the same regardless of the significance or the degree of down regulation, suggesting that the binding of down regulated genes are more or less randomly distributed. These data indicate that promoter regions of genes

bound by CYS-3 that have reduced expression under sulfur starvation may be false positives.

Conversely, we find that likelihood of CYS-3 binding correlates with degree of up regulation in sulfur starvation. We expect this as CYS-3 is a known positive regulator responsive to sulfur starvation. This is further evidence that our DAPseq peaks are reliable, although some false positives likely exist. Additionally, we can conclude that genes that show greater up regulation in sulfur starvation are more likely to be true direct targets of CYS-3.

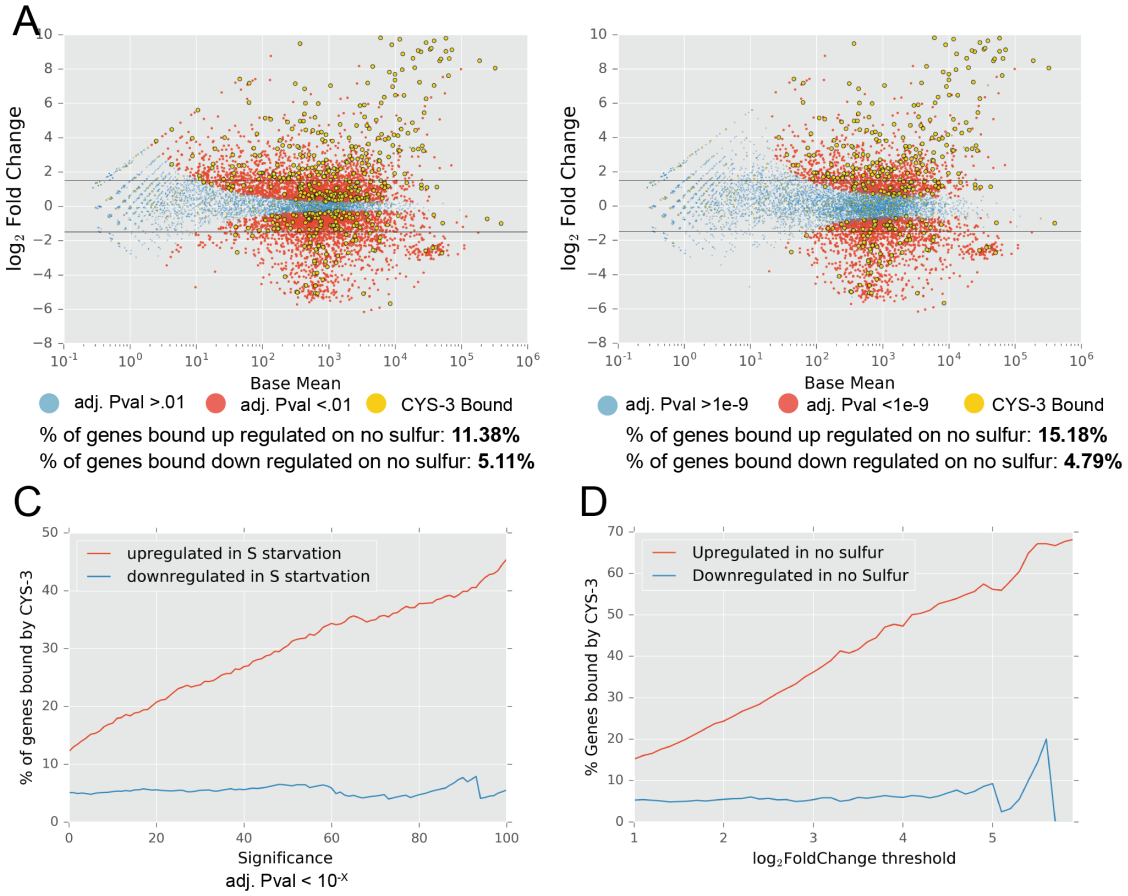


Figure 5-20 CYS-3 DAP overlayed onto sulfur starvation DEseq *A, B). Scatter plots DEseq data of high sulfur vs no sulfur where reference condition is high sulfur. Significantly differentially expressed genes are demarcated in red, while insignificant in blue, given adjusted pValue cutoffs that differ between plot A and B. Genes bound by CYS-3 are demarcated in green. C,D) line plots denoting % of genes bound as function of (C) pval significance or (D) log_2 fold change. C and D implicate CYS-3 as is a positive regulator of gene expression.*

We used the binding peaks from this list of CYS-3 genes that were also upregulated with a 4X foldchange on no sulfur vs high sulfur. These included 142 unique peaks. By taking the sequences of these binding peaks, we used MEME

v4.12.0 (184) to build a consensus motif common within all input sequences. A single motif with a significant E-value of 4.7×10^{-78} was discovered (Figure 5-21).

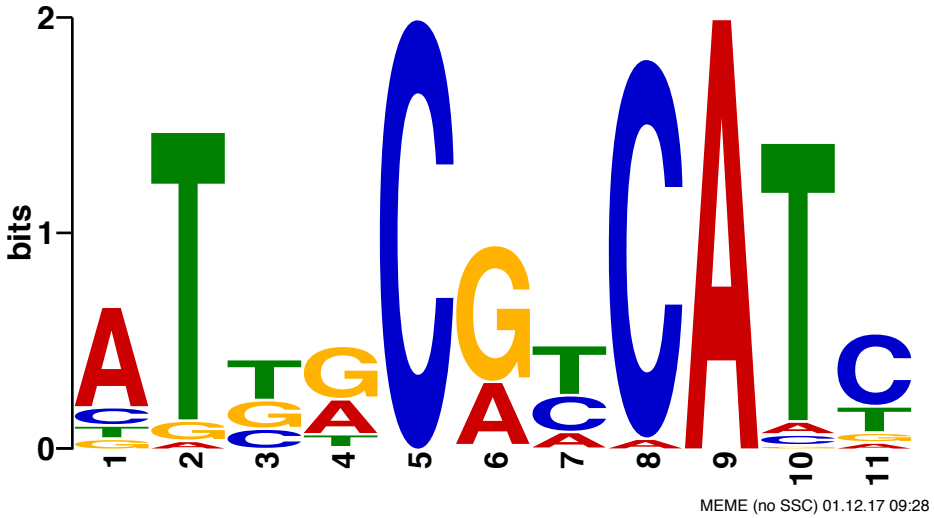


Figure 5-21 CYS-3 DNA binding motif

5.8.3. ***Conclusions and discussion***

We found that our CYS-3 DAPseq data to be in accordance with previous literature. CYS-3 has been shown to be the positive regulator for a number of critical sulfur acquisition and metabolic genes. We show that many of these genes are directly bound by CYS-3. We also found 6 additional MFS transporters with unknown function that are induced under sulfur starvation are also bound by CYS-3. These transporters likely play a role in securing additional sources of sulfur, but their target substrates remain a mystery. When we examined the portion of genes bound by CYS-3 that were also up regulated under sulfur starvation, we found a high correlation between up regulation and CYS-3 binding. All in all CYS-3 bound ~11% of all genes up regulated on sulfur starvation (adjusted Pvalue < 0.01), but significantly higher percentages up to 50-60% of genes when significance threshold or differential expression threshold were increased. The function of many of these genes still remains a mystery, but they are clearly part of the sulfur starvation response. CYS-3 also binds a number of genes down regulated on sulfur starvation. We hypothesize that these genes are likely false positives, unless CYS-3 has an additional undescribed negative regulatory role. This is supported by the general random distribution of CYS-3 binding genes within the gene set that is down regulated in sulfur starvation, which indicates that CYS-3 binding does not correlate to down regulation, and thus likely CYS-3 does not have a repressive role. Finally, we do see a number of genes up regulated on sulfur starvation and bound by CYS-3 where function is unknown. Starvation is a complex response, and additional research must be conducted to determine the role of many of these genes that are induced during sulfur starvation.

5.9. Direct binding sites of VIB-1 indicate alternate mode of action regarding the regulation of carbon metabolism than previously described

5.9.1. *Introduction*

VIB-1 is a transcription factor with a complex role in *N. crassa* biology. It was first identified in 2002 as a locus that when mutated reduced *het-c* mediated heterokaryon incompatibility. More simply, a deletion of vib-1 allowed two vegatatively incompatible strains to fuse, and also partially suppressed hyphal compartmentation and death (44). Furthermore, *vib-1* is also required for *mat* heterokaryon incompatibility, which is defined by incompatibility of strains with different mating types to fuse vegatatively (45). Thus *vib-1* is necessary for multiple types of heterokaryon incompatibility. Additionally *Δvib-1* strain showed reduced extracellular phosphatase activity (42). *XprG*, the homolog of the *vib-1* in *Aspergillus nidulans,* is involved in regulating protease production under carbon and nitrogen starvation (44, 46, 47). More recently, *vib-1* was also shown to be involved in cellulase production, with its absence leading to weak induction of *clr-2* and cellulases and dramatically reduced growth on Avicel. Two transcription factors, CRE-1 and COL-26 involved in carbon catabolite repression and glucose sensing, were hypothesized to act downstream of VIB-1 (48). These diverse phenotypes indicate that *vib-1* is involved in many aspects of *N. crassa* biology and begs the question of the interconnection of regulatory networks affected by VIB-1. We attempt to dissect the processes directly affected by VIB-1 from processes that may be modulated indirectly downstream of VIB-1 DAPseq. Our data confirms that VIB-1 is a likely positive regulator, and that it directly binds to promoters of genes involved in heterokaryon incompatibility as well as to promoters of genes encoding phosphatases, proteases, PCWDEs and *clr-2*, all of which were indicated to be downstream of *vib-1* in previous studies.

5.9.2. *Results*

General overview of VIB-1 DAPseq data

From DAPseq, we discovered 2688 promoter regions bound by halo-VIB-1 up to 3kb upstream from ATG translation initiation site. This number is approximately 25% of the annotated genes in the genome and so we expect that a number of these peaks are false positives. This large number of peaks indicates that VIB-1 is capable of binding many regions of the genome, more so than many of the other TFs that we tested with DAPseq. If we examine at only genes bound up to 1KB upstream of the ATG start site narrow in on a set of 1216 genes, removing some of the noise in the data but also possible true binding peaks.

Comparison of bound genes to RNAseq data from Δ*vib-1* versus WT indicate positive regulation of by VIB-1

We compared this set of 1216 DAPseq bound genes with genes down regulated in the Δ*vib-1* strain relative to WT on no carbon and Avicel conditions. RNAseq data was acquired from a previous study examining the role of VIB-1

regarding cellulase production (48). Authors compared WT *N. crassa* to *Δvib-1* strain under no carbon and Avicel switch experiments similar to the growth protocol for our experiments. A 100ml flask of Vogel's minimal media was inoculated with 10^8 conidia and grown for 16 hours, at which point mycelia was centrifuged, washed and transferred to 1X Vogel's with either 2% Avicel or no carbon. Mycelia was then harvested and flash frozen for RNAseq. Differential expression analysis was conducted with CuffDiff (198) between *Δvib-1* versus WT no carbon and WT 2% Avicel. When we examined the genes that were up regulated in WT under either starvation or Avicel conditions, we found that a large portion of the WT significantly up regulated genes were also bound by VIB-1 within 1KB upstream of the ATG start site (Figure 5-22 A and B). Under Avicel this was 110/232 genes or 47%. Under no carbon this was 75/199 or 37%.

To determine whether or not VIB-1 acts as a repressor or activator of genes, we looked at the percentage of genes bound by VIB-1 as we increased the threshold of $\log_2$ fold change. If VIB-1 is a positive regulator, we expect that VIB-1 bound genes have increased expression in WT compared to *Δvib-1*. We should also observe that the more reduced in expression a gene is, the more likely it is to be bound by VIB-1. Thus a plot of fraction of genes bound as a function of $\log_2$ fold change threshold should have a positive slope. Furthermore, down regulation of genes in WT should be due indirect consequences of the *vib-1* deletion or due to noise. Contrastingly, for genes up regulated on *Δvib-1* and down regulated in WT, the plot of the percentage of genes bound by $\log_2$ fold change should have a slope of 0. When we judge our data by these criteria we see that VIB-1 is likely a positive regulator rather than a repressor of gene regulation (Figure 5-22 C and D). Line graphs plotting the percentage of genes bound by VIB-1 as $\log_2$ fold change threshold is increased show a positive slope for WT up regulated genes for both no carbon and Avicel conditions, and more or less zero slope for WT down regulated genes. We also observe a much smaller number of WT down regulated genes in total in both Avicel and no carbon conditions.

To access which genes are direct targets, we selected the genes that were upregulated in WT 4X foldchange under Avicel or no carbon conditions that were also bound by VIB-1 according to DAPseq. This selection resulted in 75 direct binding partners of VIB-1 under Avicel, and 45 partners under no carbon.

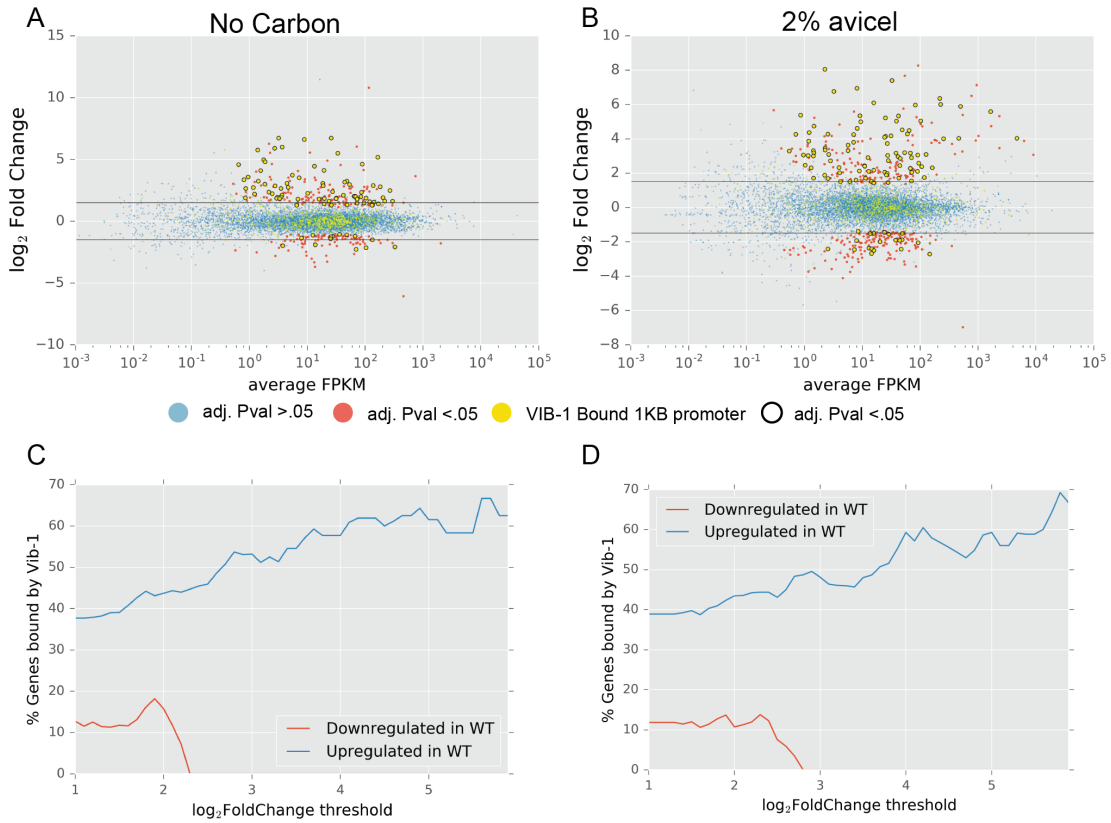


Figure 5-22 VIB-1 RNAseq with DAPseq data overlay. *A, B) Plots of differential expression analysis between WT and Δvib-1 on no carbon and Avicel respectively with Δvib-1 as the reference strain. Promoter regions bound by VIB-1 highlighted in green. C,D) line plots of % of genes bound by VIB-1 as a function of log2 fold change threshold. Positive slope of down regulated genes on Δvib-1 indicate that VIB-1 is likely a positive regulator.*

VIB-1 binding sites provide single consensus motif

We used the binding peaks from this list of VIB-1 direct targets to identify a binding motif. By taking the sequences of the binding peaks, we used MEME v4.12.0 (184) to build a consensus motif common within all input sequences. Two very similar motifs were discovered depending on whether input sequences were from Avicel or no carbon data (Figure 5-23). Both share a T separated by three base pairs followed by AC. The two base pairs towards the 5' end are biased for A's and the two bases following T are biased for GC.

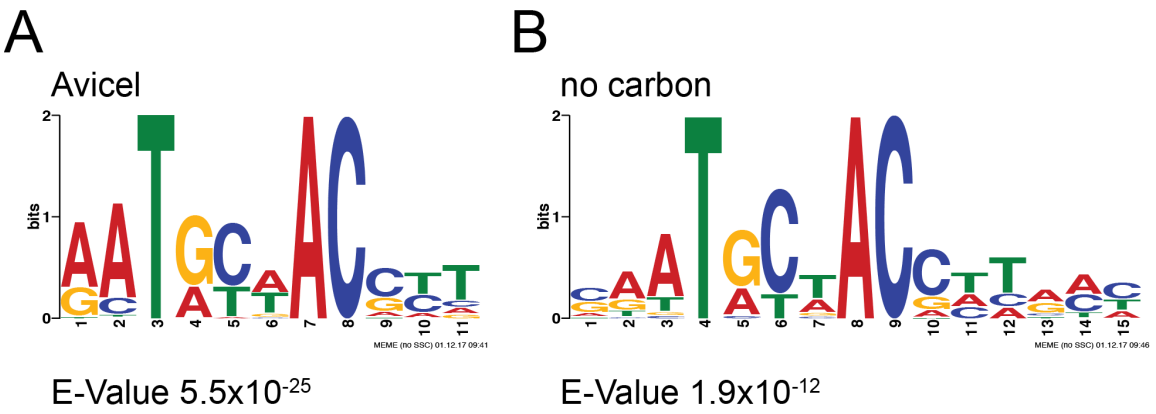


Figure 5-23 VIB-1 DNA binding motifs from Avicel and no carbon direct targets

VIB-1 DAPseq peaks in accordance with previous research regarding VIB-1 function in heterokaryon incompatibility

Previous research indicated that VIB-1 is required for expression of HET domain genes *het-6, pin-c,* and *tol*, but not *het-c* (47). A more recent study investigated all of the putative het domain genes in the *N. crassa* genome and found an additional key het domain gene *het-e*. VIB-1 DAPseq show direct binding of all 4 key het domain genes *het-6, pin-c, tol,* and *het-e*. These results along with additional het domain genes annotated from Zhao et al. 2015 (199) that show binding by VIB- are summarized in Table 5-12. Het domain genes NCU05840, *het-e,* and NCU07230 each have two binding sites in their respective promoter regions. *tol (*NCU04453) and *het-6* (NCU03533) are also down regulated in the *vib-1* strain in both no carbon and Avicel conditions (48), an observation that further confirms the authenticity of these binding sites.

Table 5-12 VIB-1 binding of HET domain genes explored in Dementhon et al. 2006 and Zhao et al. 2015 (47, 199)

Gene ID	Annotation	Fold Enrich	Distance to ATG
NCU05919	hypothetical protein	4.64	163
NCU03533	"heterokaryon incompatibility-6, het-6"	11.88	202
NCU04734	hypothetical protein	7.22	408
NCU03494	"partner for incompatibility with het-c, pin-c"	14.52	442
NCU05840	hypothetical protein	85.21	407
NCU05840	hypothetical protein	3.77	994
NCU07596	hypothetical protein	8.65	1064
NCU04835	hypothetical protein	5.55	1746
NCU09037	het-e	16.77	360
NCU09037	het-e	5.12	1912
NCU11383	hypothetical protein	8.36	2315

NCU17278	TOL	4.01	2371
NCU07230	hypothetical protein	69.64	1473
NCU07230	hypothetical protein	5.84	2767

VIB-1 direct binding can explain reduction in protease production

Deletion of *vib-1* and its homolog *xprG* in *A. nidulans*, causes a reduction in extracellular protease activity (46, 47). To determine if extracellular proteases are bound by VIB-1, we searched for all proteases, peptidases and proteinases in the *N. crassa* genome with a signal peptide as determined by SignalP (20). We found 27 genes that fit this criteria. Eight of these proteases were bound by VIB-1. This is approximately the number we would expect just by random chance, since about a quarter of all promoter regions are bound by VIB-1. However, we saw that multiple genes from this list were bound in multiple locations in their promoter sequence, thus increasing the chance that they are truly bound by VIB-1. The promoter region of metalloprotease-1 was actually bound by VIB-1 in three locations upstream of the ATG start site, and shown to be down regulated ~5 fold in the Δ*vib-1* mutant as compared to WT under Avicel conditions (48). Most of the protease binding sites are also within 1kb of the ATG start site, further supporting VIB-2 regulation. The table below summarizes the protease promoters bound by VIB-1 (Table 5-13). These results can explain reduced protease activity in the supernatants of the *Δvib-1* strain under 16 hour carbon starvation and nitrogen starvation as described previously (47). Additionally, 12 nitrogen transporters of various types were also bound by VIB-1 indicating a possible further role of VIB-1 in nitrogen scavenging (

Table 5-14).

Table 5-13 Promoters of genes encoding proteases that are bound by VIB-1

Gene ID	Annotation	Fold Enr.	Dist. ATG
NCU00673	serine protease p2	4.80	917
NCU00263	serine endopeptidase	6.01	589
NCU00263	serine endopeptidase	6.27	1779
NCU00477	carboxypeptidase Y	4.99	482
NCU06720	carboxypeptidase cpdS	5.94	1541
NCU07200	metalloprotease 1	5.58	2613
NCU07200	metalloprotease 1	5.62	1332
NCU07200	metalloprotease 1	18.86	381
NCU05071	metallo-endopeptidase	4.32	1977
NCU09992	serine peptidase	6.93	1424
NCU06055	extracellular alkaline protease	9.00	1865
NCU06055	extracellular alkaline protease	6.85	624

Table 5-14 Promoters of genes encoding nitrogen transporters bound by VIB-1

Gene_ID	Annotation	Fold Enr.	Dist. To ATG
NCU00765	amino acid permease 2	4.46	1908
NCU01481	amino acid transporter	7.08	2239
NCU07129	amino-acid permease inda1	2.80	221
NCU06613	ammonium transporter	14.91	382
NCU05843	ammonium transporter 1	3.35	2721

NCU01065	ammonium transporter MEP2	5.22	837
NCU07175	APC amino acid permease	4.70	150
NCU05168	arginine transporter	5.39	2246
NCU00758	formate/nitrite transporter	8.41	1574
NCU04435	general amino acid permease AGP3	5.12	410
NCU05830	general amino-acid permease GAP1	6.95	415
NCU01977	urea transporter	8.05	810

VIB-1 direct binding sites can explain reduction in phosphatase production

Previous literature showed that deletion of *vib-1* was associated in a reduction of extracellular phosphatases. In *A. nidulans,* loss of function mutations in *xprG,* a homolog of *vib-1,* caused reduction in acid phosphatase activity (46). In *N. crassa*, mutations in *vib-1* caused a reduction in levels of a repressible acid phosphatase, but no reduction levels of three other phosphatases tested (44). We are unable to determine the loci of these phosphatases, as they were described before the genome of *N. crassa* was sequenced. We do however see a number of phosphatases bound by VIB-1, which could explain the reduction of the acid phosphatase activity in *vib-1* mutant (Table 5-15).

We wanted to further explore the role of VIB-1 in phosphatase expression and phosphate acquisition. We selected the top genes positively regulated by NUC-1, a known positive regulator of phosphatases and phosphate transporters (Section 4.6), for initial examination. We found that VIB-1 binds 4 of the 5 top phosphate acquisition genes in the NUC-1 regulon (Table 5-15). Only the inorganic phosphate transporter *pho-5* was not bound. We searched the remaining of the VIB-1 binding sites for phosphatases and found an additional endopolyphoshphatase and a repressible alkaline phosphatase gene (Table 5-15). This data indicated that VIB-1 may directly regulate phosphate acquisition, however it is difficult to rule out binding peaks as false positives at this point in time.

Table 5-15 VIB-1 bound to promoters of genes encoding phosphatases and phosphate transporters

Gene ID	Annotation	Fold Enr.	Dist. To ATG
NCU01376	alkaline phosphatase pho-2	9.76	586
NCU08643	acid phosphatase pho-3	12.96	277
NCU09564	inorganic phosphate transporter pho-4	5.62	1942
NCU08325	inorganic phosphate transporter pho-5	N/A	N/A
NCU10038	glycerol diester phosphodiesterase	20.33	2430
NCU02996	endopolyphosphatase	4.45	744
NCU08997	repressible alkaline phosphatase	4.03	523

VIB-1 may play a role in sulfur acquisition

After taking note of the role of VIB-1 in direct regulation of nitrogen and phosphate acquisition, we reasoned that VIB-1 might also be involved in sulfur acquisition. A well-researched mechanism for sulfur acquisition has been

established in *N. crassa* that makes use of a single transcription factor CYS-3, that is repressed by SCF complex under sulfur rich conditions (147). During sulfur limiting conditions, CYS-3 is active and directly binds and up regulates a number of sulfur acquisition genes (see section 5.8). A set of sulfur acquisition genes was well characterized after the sequencing of the *N. crassa* genome, as we use this list to conduct an initial examination into whether sulfur acquisition may be affected by VIB-1. We found that promoters of 9 of the 20 sulfur acquisition genes were bound, including key genes like arylsulfatase and both sulfate permeases. These three genes are three of the most highly up regulated genes under sulfur starvation, indicating their importance for in sulfur acquisition. Additionally Sulfate permease II was bound in two places within the promoter region, indicating increased likelihood that it is a true positive. Results regarding VIB-1 binding of sulfur acquisition gene promoters is summarized in the table below (Table 5-16). Experiments of *Δvib-1* in comparison to WT in sulfur limiting conditions must be conducted to confirm these results.

Gene ID	Annotation	Fold Enr.	Dist. To ATG
NCU06041	Arylsulfatase (ARS-1)	5.71	371
NCU05340	Alkanesulfonate monooxygenase	3.76	2841
NCU06112	Cysteic acid decarboxylase	4.98	2057
NCU09738	Taurine dioxygenase	5.49	672
NCU06625	Cysteine dioxygenase	4.59	1099
NCU06931	Sulfite oxidase	5.23	909
NCU03235	Sulfate permease I (CYS-13)	3.83	1278
NCU04433	sulfate permease II (CYS-14)	4.59	2871
NCU04433	sulfate permease II (CYS-14)	3.30	83
NCU04942	Methionine Permease	4.68	259

Table 5-16 Promoters of sulfur acquisition genes that show enrichment for binding by VIB-1

VIB-1 may act directly through CLR-2 and cellulases rather than CRE-1 and COL-26 to induce production of cellulases.

Previous literature showed that deletion of *vib-1* caused dramatically reduced growth on Avicel and moderately reduced growth on xylan (48). No growth defects were detected when the deletion strain was grown in media containing sucrose, cellobiose, or xylose indicating defect was caused by the lack of extracellular PCWDEs. This conclusion was further supported by reduced cellulase activity in the supernatant of *Δvib-1* strain (48). Another major finding of this study was that the double deletion of *cre-1* and *col-26,* TFs involved in glucose sensing and metabolism could rescue *Δvib-1* defect on growth under 2% Avicel .02% 2-deoxy glucose conditions. This result led authors to believe that *vib-1* acts upstream of glucose sensing and CCR, and that VIB-1 was acting upstream of COL-26, CRE-1 pathway (48). From our DAPseq data, we propose that VIB-1 acts directly on CLR-2 and cellulases rather than acting upstream of CRE-1 and COL-26. We suspect that

deletion of *cre-1* and *col-26* can boost expression of *clr-2* and cellulases enough to restore growth of the *Δvib-1* mutant on Avicel in a manner similar to how deletion of the three β-glucosidase genes (Δ3BG) can restore cellulase production in the *Δvib-1* on cellobiose (48).

Our DAPseq data suggests that VIB-1 directly regulates CLR-2, thus explaining the majority of genes down regulated in *Δvib-1* under Avicel conditions. Xiong et al. 2014 showed that clr-2 is down regulated in the *Δvib-1* approximately 5.3 fold (48). In WT strain on Avicel, *clr-2* has an average expression of ~178 FPKM, while *Δvib-1* is at 38 FPKMs. DAPseq data shows that *clr-2* promoter is bound by VIB-1 538bp from the ATG with a ~3.8 fold enrichment. Scatter plots of our differential expression analysis results of WT vs *Δvib-1* mutant, shows that many of the up regulated genes in WT are indeed bound by VIB-1 (Figure 5-22B,D). When we overlay genes within the clr-2 regulon, we see that many of the remaining genes that are up regulated in WT are in fact a part of the CLR-2 regulon (Figure 5-24). A number of these genes are actually both bound by VIB-1 and CLR-2 and are highlighted in green squares in this plot. In total, there are 232 genes up regulated in WT compared to *Δvib-1* under Avicel conditions. 110 of these genes are directly bound by vib-1 and 35 additional genes are part of the CLR-2 regulon and not bound by VIB-1. 145/232 or 62% of these genes can therefore be explained by VIB-1 direct binding or reduced CLR-2 caused by deletion of *vib-1.* Furthermore, the majority of the remaining genes that show increase expression in WT are differentially expressed below 8X fold change (Figure 5-24). All in all, these data indicate that the majority of highly differentially expressed genes in the *Δvib-1* mutant are likely due to direct regulation by VIB-1 protein on the gene or on *clr-2* promoter.

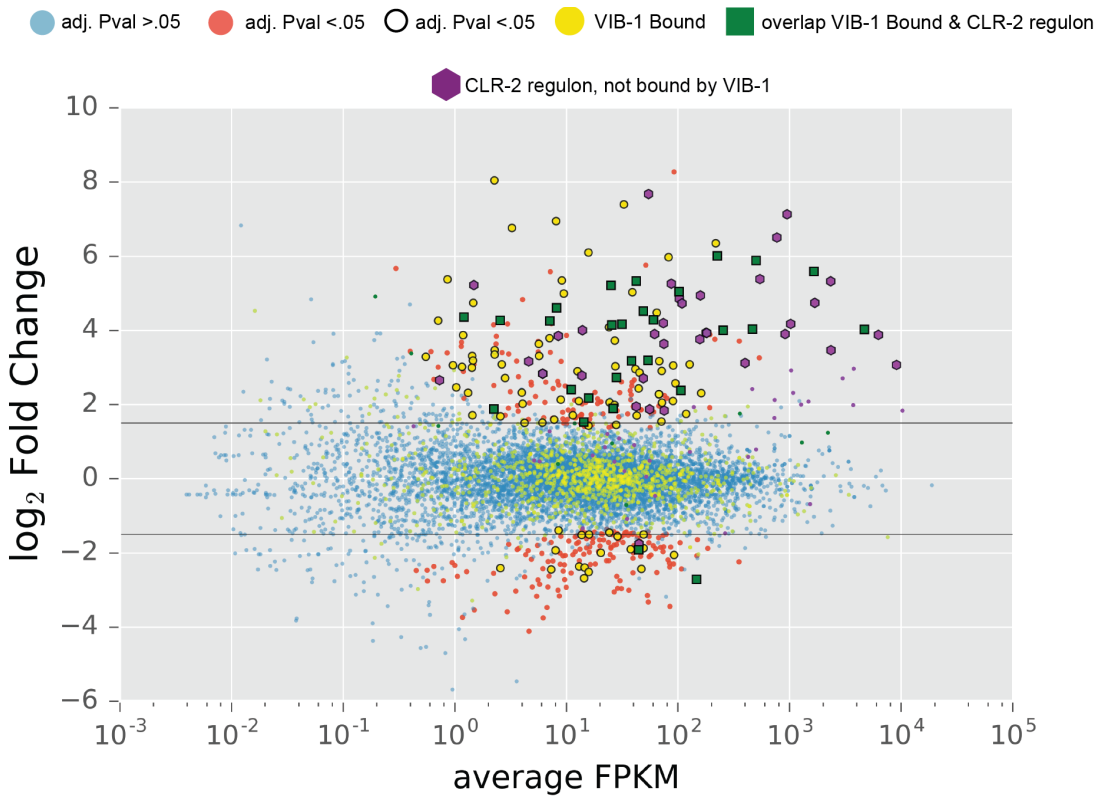


Figure 5-24 *Δvib-1* mutant vs WT differential expression on Avicel with VIB-1 DAPseq bound genes and CLR-2 regulon genes highlighted. *Scatter plot of Cuffdiff differential expression analysis between Δvib-1 and WT on Avicel from Xiong et al. 2014* (48) *with Δvib-1 as the reference condition. VIB-1 DAPseq bound genes within 1kb of ATG start site are highlighted in yellow. Genes that appear in the CLR-2 regulon and bound by VIB-1 are highlighted in green boxes. Genes in the CLR-2 regulon only are highlighted in purple hexagons.*

VIB-1 binds additional TFs involved in PCWDE production and alternative sugar metabolism.

VIB-1 binding sites were also identified in the promoter regions of a number of other important transcription factors responsible for regulating hemicellulose and pectin utilization, which can further explain the set of genes down regulated in Δ*vib-1* on Avicel. We summarize the binding peaks in Table 5-17. TFs containing N/A are not bound by VIB-1.

Table 5-17 Promoter regions of nutrient metabolism TFs bound by VIB-1

Gene ID	Annotation	Fold Enr.	Dist. To ATG	Fold change WTvsΔvib1 Avicel
NCU07705	clr-1	N/A	N/A	-0.79
NCU08042	clr-2	3.81171	538	-5.68
NCU06971	xlr-1	5.98853	1375	-1.59
NCU09033	pdr-1	N/A	N/A	-0.11
NCU04295	pdr-2	4.53641	589	-0.73
NCU05414	ara-1	6.59157	2168	-0.74

NCU01074	hem-1	5.32921	2113	-0.31
NCU00445	man-1	N/A	N/A	0.00
NCU09068	nit-2	N/A	N/A	-0.02
NCU08294	nit-4	N/A	N/A	0.00
NCU09315	nuc-1	N/A	N/A	-1.00
NCU03536	cys-3	N/A	N/A	0.00

According to our DAPseq results, additional TFs *xlr-1, pdr-2, ara-1, and hem-1* are all bound by VIB-1. We have no RNAseq to confirm that these binding sites are true positives, but some evidence indicates that the *xlr-1* binding site is authentic and functional. A previous study indicated that Avicel is likely contaminated with some xylans (43), as xylanases are shown to be expressed under Avicel but not induced by cellobiose. We observed the same results in our data (Figure 3-5), where 2mM cellobiose was unable to induce the expression of any of the xylanases observed on 1% Avicel. We also know that xylanases are mainly under positive regulation by XLR-1, which is activated by xylan (126). With this knowledge present, combined with evidence that *xlr-1* is bound by VIB-1 and down regulated in a Δ*vib-1* mutant, we hypothesize that part of the XLR-1 regulon should be down regulated in the Δ*vib-1* strain independent of VIB-1 direct binding. When we overlaid XLR-1 regulon onto our WT vs *Δvib-1* Avicel differential expression data, we see that a number of genes down regulated in Δ*vib-1* that cannot be explained by VIB-1 binding or CLR-2, can be explained by the reduced XLR-1 activity (Figure 5-25). Thus we conclude that genes up regulated in WT in comparison to *Δvib-1* that are not directly bound by VIB-1, are up regulated due to the activity of CLR-2 or XLR-1 not present in the *Δvib-1* mutant.

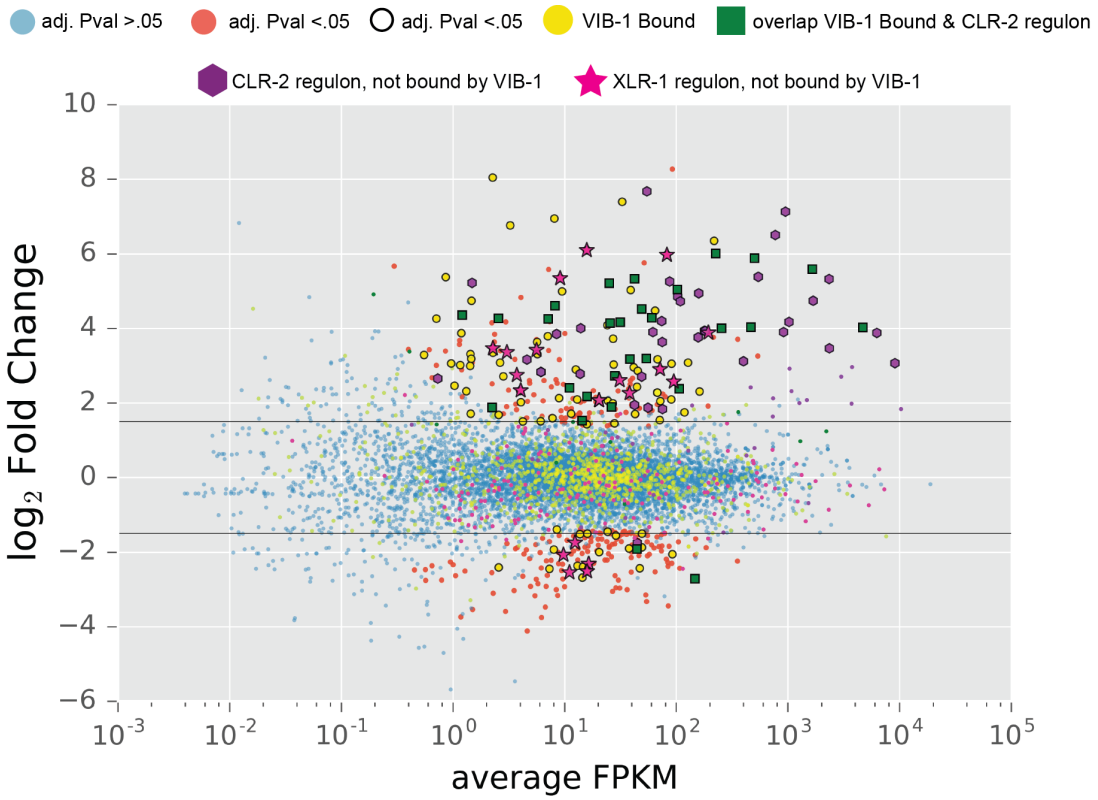


Figure 5-25 The Δvib-1 vs WT differential expression data overlaid with VIB-1 bound and CLR-2 regulon highlighted, AND non overlapping XLR-1 regulon highlighted. *Scatter plot of Cuffdiff differential expression analysis between Δvib-1 and WT on Avicel from Xiong et al. 2014* (48) *with Δvib-1 as the reference condition. VIB-1 DAPseq bound genes within 1kb of ATG start site are highlighted in yellow. Genes that appear in the CLR-2 regulon and bound by VIB-1 are highlighted in green boxes. Genes in the CLR-2 regulon only are highlighted in purple hexagons. Genes in XLR-1 regulon are highlighted in magenta stars.*

5.9.3. ***Summary and Discussion***

In this section, we show that VIB-1 is a strong positive transcription factor that plays a role in many different cell processes. Most of these processes are related to nutrient acquisition, but also cell-cell fusion and heterokaryon incompatibility. Although VIB-1 has never been observed in a direct positive regulatory role, we show here that genes differentially expressed in the *Δvib-1* strain have an increasing likelihood of being bound by VIB-1 with increasing down regulation in *Δvib-1* mutant as compared to the WT strain. This pattern is characteristic of positive regulators and observed with CYS-3, XLR-1, CLR-2, NIT-2 and other positive regulators discussed in previous sections. We further show that many HET domain genes, proteases, phosphatases and cellulase promoters are all bound directly by VIB-1. These data explain the phenotypes observed in Δ*vib-1* strain in previous research: no expression of *het-6, pin-c and tol* (47), reduced production of proteases and phosphatases (44, 46), and reduced expression of cellulases (48).

We further showed that TFs important for plant cell wall utilization including *clr-2*, *xlr-1, pdr-2, ara-1* and *hem-1* and are all bound by VIB-1. This data suggests that VIB-1 is a master regulator of these TFs. We have no strong evidence that hemicellulose and pectin regulators (*xlr-1, pdr-2, ara-1* and *hem-1)* are indeed within the VIB-1 regulon. However, RNAseq data from a previous study (48) combined with our direct binding data for VIB-1 indicates that *clr-2* is bound by and requires VIB-1 for full expression. It would therefore not be surprising if VIB-1 is required for full expression of the hemicellulase and pectin regulators under their inducing conditions. It has already been shown that Δ*vib-1* has reduced growth on xylan (48), an additional hint that VIB-1 may be a positive regulator of *xlr-1*. An RNAseq experiment comparing WT to Δvib-1 on xylan could easily test this hypothesis, as we would expect to see expression of *xlr-1* and VIB-1 bound xylanases to be down regulated in the Δ*vib-1* strain. If a *Δvib-1* reduced growth phenotype is present in pectin, similar experiments could be conducted to validate direct binding of VIB-1 to *pdr-2* and *ara-1* promoters.

Our findings and along with past findings are consistent with a model where VIB-1 is a general starvation response transcription factor, or perhaps a transcription factor that allows for basal expression of nutrient acquisition genes. The most basic evidence for this model are our observations that VIB-1 is bound to the promoters of a diverse set of genes involved in nutrient acquisition. We use the role of VIB-1 in cellulase production as model for how we believe VIB-1 behaves with regard to nutrient acquisition of other types (nitrogen, sulfur, phosphate). Previous to this work, Coradetti et al., 2012 and others (27, 183)established a model for how cellulase secretion is regulated in *N. crassa:* Interaction between cellobiose and CLR-3 activates CLR-1, which directly binds and promotes transcription of *clr-2*, whose protein product positively regulates the majority of *N. crassa's* cellulases. However, in order for there to be cellobiose present initially, basal expression of cellulases is required. This allows enough cellobiose signal to activate CLR-1 and in turn full transcription of cellullases. We believe that VIB-1 is the transcription factor providing this basal transcription of cellulases through its direct promotion *clr-2* in combination with a small set of cellullases. If the cell does encounter cellulose, cellobiose will begin to trickle into the cell and initiate a positive feedback loop of CLR-1 activation further *clr-2* expression, and increasing cellobiose signal until full expression of cellulases occur. Without VIB-1, this positive feedback loop is not fully initiated, and full expression of CLR-2 and the cellulases required for normal growth on Avicel are not achieved. This can explain the inability of the *Δvib-1* strain to grow on Avicel in observed in Xiong et al., 2014 (48).

This model is contradictory to what authors concluded in Xiong et al., 2014 (48) mainly due to results examining growth phenotypes of *Δvib-1* in combination with deletion of *cre-1* and *col-26* under 1% Avicel and 0.2% 2-deoxyglucose (2DG). 2DG is an non-metabolizable analog of glucose that activates CCR due to the false signal of high glucose levels. Thus with 2DG present, cellulases are repressed. COL-26 seems to be involved in glucose sensing and relaying that information to CRE-1. Experiments showed that deletion of *col-26* and *cre-1* together rescued the growth phenotype of *Δvib-1* on Avicel and 0.2% 2DG. From this data, authors assumed that CRE-1 and COL-26 are downstream of VIB-1. Our DAPseq data does not support this

hypothesis. Neither *cre-1* nor *col-26* are bound by VIB-1, nor do we see VIB-1 acting as a repressor. We hypothesize that in the *Δcre-1 Δcol-26 Δvib-1* triple deletion strain, CRE-1 mediated repression is completely off. CRE-1 target CDT's are no longer repressed, and enough cellobiose can be imported into the cell for activation of CLR-1 positive feedback cycle. Activation of CLR-1 can be maintained at high levels as expression of the cellodextrin transporters (CDT)s are not turned down even though glucose is present at high concentrations in the cell. This in turn induces expression of *clr-2* to levels that normally require VIB-1 presence. This explanation is in accordance with our data and previous data from Xiong et al. 2014.

5.10. **Summary and discussion**

We found DAPseq to be a successful way to assay direct binding of fungal TFs. We showed that DAPseq results were consistent with ChIPseq results, but that neither are perfect tools. Both ChIPseq and DAPseq include many false positive peaks. Binding at these sites may be true, but probably are not biologically relevant. For instance peaks in the middle of genes or centromeric regions may be real but are likely not functioning to promote expression of any genes. Many of the peaks within 3KB upstream of ATG start sites may also be false positives. RNAseq was helpful in filtering out these false positive peaks. For TFs with complementary RNAseq data, we combined RNAseq and DAPseq to create a final set of genes likely directly regulated by the TFs we tested.

We were able to expand the regulons of CYS-3 and NIT-2. These transcription factors have been studied frequently regarding their core regulons and interaction partners. Their full regulons had not been explored until now. We see that CYS-3 and NIT-2, the major regulators of sulfur and nitrogen metabolism respectively, have much wider influence on gene regulation than previously thought. CYS-3 directly regulates a number of previously untested genes such as taurine dioxygenases, and undescribed MFS transporters likely involved in sulfur acquisition. Many genes bound by CYS-3 and up regulated under sulfur starvation have yet to be explored. Our NIT-2 data also uncovered a number of undescribed targets of NIT-2. A number of transporters that act on various nitrogen sources, and a few peptidases are both bound and down regulated in *Δnit-2* mutant under nitrogen starvation conditions.

Our data on CRE-1 and VIB-1 provides interesting details regarding carbon acquisition and sensing in *N. crassa.* Both of these TFs are well conserved in the fungal kingdom so such findings may be relevant to the biology of many fungi. We believe that VIB-1 it is responsible for the basal transcription of cellulases that allows cellulose to be detected outside of the cell. It is also required for full expression of *clr-2* and cellulases when cellulose is the major carbon source. We believe these statements to be true because of four lines of reasoning. The first is that VIB-1 is active in almost all conditions. This is evidenced by its transcription profile data. We observe that *vib-1* is always expressed at least 40 FPKM in all the conditions we have data on. Data from previous studies show that either a growth or expression phenotype on all conditions tested. This includes sucrose (47), phosphate and nitrogen starvation (44, 46) and cellulose conditions (48), thus

indicating VIB-1 is active on all these conditions. Furthermore, the only time VIB-1 has been observed absent from the nuclei is during development of conidia (47). Secondly, VIB-1 directly regulates *clr-2* and cellulases as shown by data in this section. Thirdly, deletion of the *vib-1* prevents it from growing on cellulose (48). This informs us that VIB-1 is critical for either sensing cellulose, expressing cellulases or both. Fourth, *vib-1* transcription is not under CCR, nor are its direct binding targets. We have no data indicating that *vib-1* is bound by CRE-1, nor does *vib-1* have reduced expression on sucrose a symptom characteristic of CCR mediated genes. These lines of reasoning indicate that VIB-1 should be promoting expression of constitutively active CLR-2, and some cellulases in all conditions shown so far except for conidial development.

CRE-1 DAPseq indicates that CRE-1 mode of action may be different from previously hypothesized. Past research has shown that CRE-1 is bound to the promoters of both *xlr-1* and xylanases in *A. nidulans.* Research in *N. crassa* showed that CRE-1 binds a number of cellulases and glucose transporters through ChIP-PCR (190). Our DAPseq data is contradictory to much of this data, as we only find 2 of 9 genes overlapping with ChIP PCR data. We were surprised that very few carbon catabolic TFs were bound by CRE-1; only *clr-1* and *nit-4* were bound. Instead, we observed that a major portion of the promoters of predicted sugar transporters were bound by CRE-1, including all three CDTs. If our data is indeed correct, this would indicate CRE-1 mediated CCR acts strongly through transporters that bring signal transducing sugars into the cell that are ultimately responsible for the up regulation of genes encoding PCWDEs.

With the combined CRE-1 and VIB-1 data, we propose a revised model for cellulase regulation in *N. crassa* (Figure 5-26). When cellulose is present and glucose absent, VIB-1 promotes expression of *clr-2* along with a small set of PCWDEs, so that minute amounts of cellulases are secreted. These cellulases cleave cellulose and cellobiose is transported into the cell by the sugar transporters CDT-1 and CDT-2. Cellobiose (or a modified version of cellobiose) inactivates the repressor CLR-3, which allows activation of CLR-1, which is present in low levels in the cell. At this point, CLR-1 promotes expression of *clr-2* and cellulases, thus beginning a positive feedback loop until full expression of cellulase genes occurs. As the cell is flooded with cellobiose, it is cleaved into glucose. Glucose induces CRE-1 mediated CCR, which damps down expression of *clr-1* and the cellodextrin transporters *cdt-1, cdt-2* and *cdt-3.* Cellobiose uptake is reduced, and so are CLR-1 levels and activity, thus negatively regulating cellulase expression. Cellobiose, CLR-1 activity and CRE-1 are always in balance thus keeping cellulase expression steady. These data are consistent with growth defect of *Δvib-1* mutant on cellulose. With no VIB-1, low levels of cellulases are not present to initiate the positive feedback loop that allows full expression of cellulase genes. It is also consistent with rescuing of cellulase expression of *Δvib-1* with triple β-glucosidase deletion (Δ3BG) (48). With a buildup of cellobiose in the cell due to Δ3BG, CLR-1 remains active, thus promoting expression of *clr-2* and cellulases. Previous authors concluded from this data that VIB-1 acted upstream of cellobiose induction of CLR-1, while we believe this shows that VIB-1 acts independently of cellobiose and CLR-1 induction. This is a conclusion supported by our direct binding data, which shows VIB-1 acting on *clr-2*.

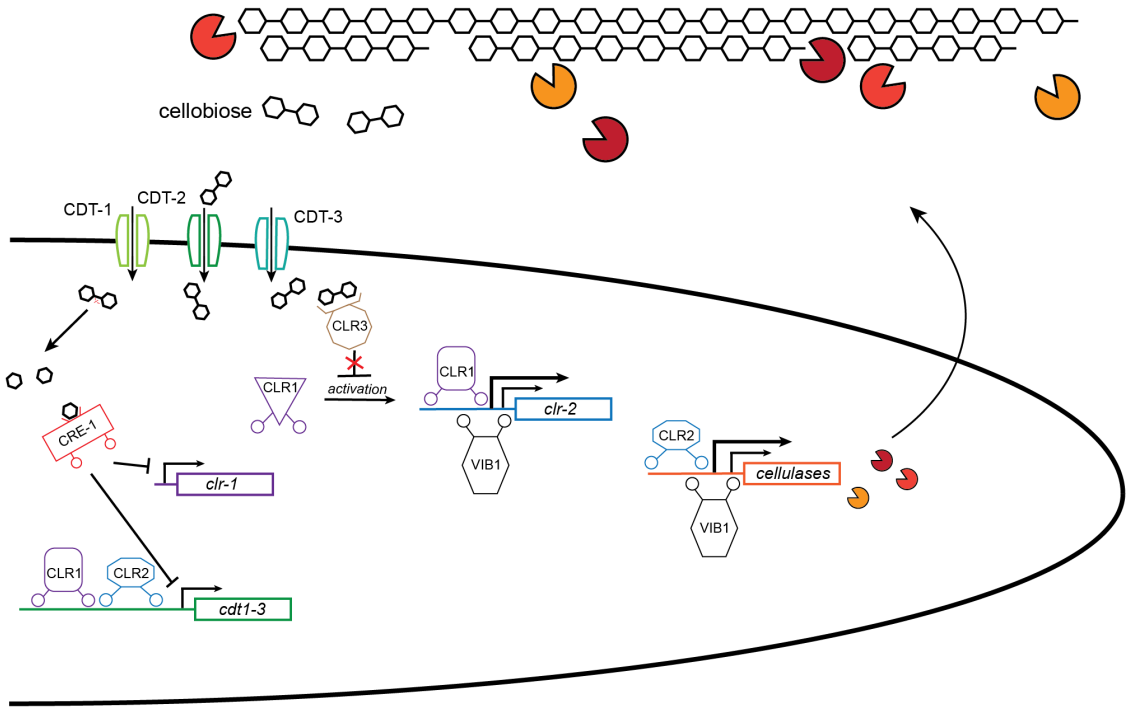


Figure 5-26 Cellulase production model with VIB-1 and CRE-1 *VIB-1 induces expression of clr-2 and cellulases for nutrient sensing, and allows full expression of cellulases under inducing conditions. Cellobiose import into the cell is turned into glucose, which represses clr-1 and cdts through action of CRE-1, which reduces expression and activation of CLR-1, thus reducing cellulase expression.*

VIB-1 may play a similar role in hemicellulase and pectinase expression as it does for cellulase expression. Major regulators of xylan, pectin and arabinan catabolism *xlr-1, hem-1, pdr-2* and *ara-1* all have VIB-1 binding peaks in their promoters. There is also evidence that *xlr-1* peak is authentic, as expression of *xlr-1* is reduced in *Δvib-1* under Avicel conditions, which contains a low level of hemicellulose contamination (48). This also fits with the idea that VIB-1 is involved in nutrient acquisition as evidenced by its regulation of proteases and sulfatases, phosphatases and transporters for these nutrients.

VIB-1 regulation of HET domain genes may be part of its role as a nutrient acquisition TF. HET domain genes allow fungi to distinguish between self and non-self cells, and initiate programmed cell death of non-self fusion pairs (199). Associated with filamentous fungi as they face nutrient limitation is the increasing frequency of fusion. It has been established that in *N. crassa* that hyphal fusion is required for efficient nutrient translocation through the colony (200). This observation indicates that nutrient limited cells have reason to increase fusion such that limited resources can be easily distributed. Indeed *Botrytis cinerea* germlings have higher rates of fusion on minimal media than in rich media (201). *Colletotrichum lindemuthianum* germlings fuse at much higher rates when completely starved (202) and *N. crassa* germling fusion is inhibited by tryptophan (203). These data lead us to believe that nutrient-limited hyphal cells may also have

increased fusion rates as well. If we assume this to be true, it is also crucial for cells to ensure that they are not fusing with the wrong partners. This might explain why VIB-1, seemingly involved in nutrient acquisition, might also play a role in HET domain gene expression. Previous research shows that HET domain genes *pin-c, het-6* and *tol* are all down regulated in the *Δvib-1* strain (47). We show that VIB-1 binds directly to the promoters of these genes. Perhaps VIB-1 increases expression of these HET domain genes to ensure fusion does not occur with non-self cells especially in the nutrient limiting conditions where there are increased rates of self fusion.

Chapter 6. **Discussion and conclusions**

A complete picture: genome to transcriptome to regulome.

This dissertation explored many aspects of *N. crassa* response to the plant cell wall material. It was an amazing opportunity to provide a complete picture of plant cell wall degradation within a single species of a filamentous fungus. We explored a diverse set of input signals from sugars, polysaccharides and plant biomass to N, P and S starvation in order to explore all the processes that modulate plant cell wall deconstruction. We discovered an intricate set of primary PCWDE expression modulating transcription factors that integrate upstream signals from a variety of sources.

Genome level information provided us with all the tools *N. crassa* has in its inventory to degrade the plant cell wall. These tools mostly take the form of hydrolytic enzymes capable of cleaving polysaccharides from the plant cell wall. Recently non-hydrolytic, lytic polysaccharide monooxygenases LPMOs have also been characterized (204, 205). Their importance in plant biomass depolymerization are finally being fully understood (206). With knowledge accumulated regarding the enzymes required for breakdown of the plant cell wall, we can assess the enzymes present in the *N. crassa* genome. We see that the *N. crassa* inventory is dominated by cellulases and cellulolytic LPMOs. Approximately 40% of PCWDEs are for cellulose degradation. This is surprising as cellulose is structurally very simple, composed only of glucose monomers. Xylanases are the second most abundant, followed by pectinases, and then various hemicelluloytic enzymes required for cleavage of sugars like arabinose galactose, fucose and mannose from parent polysaccharides (for more see section 1.4). *N. crassa* also has a number of starch degrading enzymes like α-amylases and a recently discovered starch active LPMO (19).

From the genome sequences for these PCWDEs we can also predict whether they are modified with pyroglutamate (pGlu). We show that pGlu is an important modification of proteins for increased structural stability and resistance to proteases, and can be predicted with a simple assessment of the where the signal peptide for a given protein is cleaved. This prediction is made more robust by the understanding of where the cleavage must occur in the cell. We showed that it must occur inside the ER, and that proteins with exposed N-terminal glutamine are likely to be modified with pGlu.

Transcriptome information allows us to distinguish the importance plant cell wall degrading enzymes. At the most basic level, we can appreciate which genes have low, moderate and high expression across all of our carbon conditions. PCWDEs can have extremely high expression in the tens of thousands of FPKMS. Others are expressed in the tens of FPKMs. Total abundance provides us with some information to gauge importance. The fold change in transcript level between a control condition like starvation and an inducing condition like Avicel can provide additional information. For instance, PCWDEs with the highest differential expression between starvation and Avicel are probably important for cellulose degradation (see section 3.3.3 for more information). We use this information to rank PCWDEs for their importance. This was especially useful for starch enzymes, as we observed that only a few starch enzymes are both highly expressed and differentially expressed on starch vs starvation. We also observed that certain hemicellulases, pectinases and cellulases have very low expression across all conditions, indicating their probable lack of importance.

Genome and transcriptome data also helped fill in large gaps in the understanding in sugar transport. Unlike PCWDEs, many transporters remain un-annotated. We showed that all of the known sugar transporters all belong to the same clade within the MFS superfamily. Many of the un-annotated transporters in this clade are also highly up regulated across various carbon conditions. These observations are further evidence for their involvement in sugar transport.

Expression of the vast majority of PCWDEs can be explained by primary PCWDE regulators found and described in Chapter 4. Previous to work conducted in this dissertation, clear knowledge gaps regarding pectinase and hemicellulase regulation were present. At this point in time, an extensive set of pectinase genes was annotated; yet the regulation behind the expression of these genes had not been explored. This was also true for key GalA, rhamnose and arabinose metabolic genes. These sugars are abundant in the plant cell wall and are the most abundant sugars after cellobiose and xylose (166). Mannose is another common sugar, but the presence of a single β-mannosidase indicates that *N. crassa* likely does not rely on this polysaccharide. The search was on for regulators of pectinases and the metabolism of these pectic sugars. Our collaborators, Philipp Benz and colleagues, had already found a pectinase regulator *pdr-1* through a screen of pectin growth mutants (35). This turned out the be the homolog of *RhaR* in *Aspergillus* (36). We found two additional regulators *pdr-2* and *ara-1* both of which were published on in related filamentous fungi as we worked to characterize them (37, 41). These TFs filled in the knowledge gaps associated with the regulation of pectinases and arabinose, GalA and rhamnose catabolism. We found an additional TF, *man-1,* a conserved Zn2Cys6 TF that we show regulates the single β-mannanase in the *N. crassa* genome. It may be a vestige of mannanase and mannose catabolism regulator in the evolutionary past. Perhaps this regulator can be further explored in organisms capable of more robust utilization of mannan. In summary, we found that the discovered regulators could explain expression of all PCWDEs in the *N. crassa* genome. They also explained the metabolism of the most abundant sugars in the plant cell wall. These data indicated to us that major primary regulators for the degradation of plant polysaccharides had all been found.

We used DAPseq to try and confirm direct regulation of PCWDEs of the TFs we identified. However, the protocol only worked for a minority of TFs tested, none included the new TFs described in this dissertation. We showed that the direct regulators we found modulated expression of PCWDEs through RNAseq, but could not show that these findings were explained by the direct binding of a TF to a target PCWDE. It is still possible, albeit unlikely, that the regulators we found modulate an unknown TF that then directly regulates PCWDEs. We do not think this is the case because we observed no other TFs with transcription profiles specific for pectic conditions.

An effort to include many controls in our DAPseq experiments led to additional knowledge on well researched transcription factors CYS-3 and NIT-2. We showed that these transcription factors directly regulate additional genes that are involved in sulfur and nitrogen acquisition that had yet to be connected with CYS-3 and NIT-2.

Sugars play an underappreciated role in PCWDE regulation

Although DAPseq failed for a majority of the TFs tested, the technique was successful for seven of the TFs. Perhaps protein folding worked for these TFs or perhaps these TFs require no additional partners for binding. It is also possible that these TFs expressed better in the transcription-translation reaction. Regardless of why these libraries were successful, direct binding data for VIB-1 was present in the promoters of many primary PCWDE promoting TFs. These include, *clr-2, xlr-1, pdr-2, ara-1* and *hem-1.* Researchers in our field and our lab along with myself have spent many years searching for and trying to understand the function of these primary TFs (26–29, 31, 35, 37, 38, 41, 126, 171, 183). Discovering how VIB-1 directly affects expression of these primary TFs together has been an exciting process, and changes how we view regulation of PCWDE expression in filamentous fungi. So far, we have confirmed VIB-1 regulation of *clr-2* using published RNAseq data, but additional experiments are required to confirm direct regulation of *xlr-1, pdr-2, ara-2* and *hem-1.* Our findings also fit together with previous research concluding the role of VIB-1 in nutrient acquisition (46, 47), as well as our CRE-1 binding data.

DAPseq data from carbon catabolite repression TF CRE-1 has made us rethink how CCR occurs. The new model fits surprisingly well in relation to the VIB-1 model. We found that CRE-1 binding sites are highly enriched in the promoters for genes encoding sugar transporters. CRE-1 is also bound to PCWDEs but not at the frequency observed for transporters. Surprisingly, CRE-1 is only bound to a single primary PCWDE regulator, *clr-1.* These findings indicate that VIB-1 may not be restricted by CRE-1, and that the two TFs work most effectively on separate parts of the cellulase production pathway. In our model, this is critical for VIB-1 function because VIB-1 requires cellulases and *clr-2* to remain derepressed under CCR. This way VIB-1 can still promote enough expression of cellulase genes to detect cellulose if it is present.

If VIB-1 binding sites on *xlr-1*, *pdr-2*, *ara-1*, and *hem-1* promoters are authentic, we can expand this model to include hemicellulose and pectin catabolism.

Under normal conditions, we can imagine VIB-1 to be active and promoting expression of these TFs. Thus, enough TF is around to respond to any incoming signaling sugars. Conveniently for our model, VIB-1 is also bound to all the necessary PCWDE genes that cleave signaling sugars from a parent polymer. These include pectin methylesterase, exogalacturonanase, α-arabinofuranosidases, rhamnogalacturonan lyase, rhamnogalacturonan acetyl esterase, β-xylosidase, and xylanases all bound within the first 1kb of their promoter region. These data are summarized in the table below (Table 6-1). If our hypothesis is correct, VIB-1 alone can promote expression of primary PCWDE regulators as well as provide the signaling sugars that activate these TFs when a polysaccharide is present outside of the cell.

Table 6-1 PCWDEs within 1kb of ATG bound by VIB-1 *Data from DAPseq of halo-VIB-1*

Gene ID	Annotation	Fold_Enrich	Distance_ to_ATG
NCU09041	xylulose reductase	4.54	43
NCU10045	pectin methyl esterase	2.73	90
NCU06961	exogalacturonanase	6.22	136
NCU05429	alpha-amylase	5.34	161
NCU09702	endo-beta-1,6-galactanase	12.85	187
NCU05882	endo-beta-1,6-galactanase	4.81	218
NCU07351	alpha-glucuronidase	4.06	222
NCU01867	polysaccharide monooxygenase 1 (PMO1)	3.85	222
NCU09170	alpha-arabinofuranosidase	18.58	230
NCU02240	polysaccharide monooxygenase 2 (PMO2)	8.06	249
NCU06861	glycosyl hydrolase family 43 protein	16.76	263
NCU09976	rhamnogalacturonan acetyl esterase	6.76	281
NCU09923	extracellular beta-xylosidase	8.62	304
NCU09775	alpha-arabinofuranosidase	5.11	312
NCU05598	rhamnogalacturonan lyase	15.83	345
NCU02855	beta-1,4-endoxylanase	15.10	356
NCU07190	exo-beta-1,4-glucanase or cellobiohydrolase	7.23	362
NCU02344	polysaccharide monooxygenase 1 (PMO1)	5.19	363
NCU09805	alpha-amylase	4.12	398
NCU09764	polysaccharide monooxygenase 1 (PMO1)	32.44	399
NCU02343	alpha-arabinofuranosidase	4.64	420
NCU05121	beta-1,4-endoglucanase	6.10	436
NCU09924	exoarabinanase	8.62	446
NCU04494	acetyl xylan esterase	9.14	447
NCU07974	polysaccharide monooxygenase 3 (PMO3)	12.92	502
NCU05969	polysaccharide monooxygenase 3 (PMO3)	3.56	509

NCU08746	starch binding domain-containing protein	9.12	529
NCU05037	L- and L-KDR aldolase LRA4	11.57	610
NCU09491	feruloyl esterase	11.44	682
NCU02583	extracellular alpha-glucosidase	8.40	781
NCU00988	GalA transporter Gat-1	6.42	829
NCU00985	extracellular beta-mannosidase	12.12	940
NCU08054	intracellular beta-glucosidase	6.86	962

We showed that the fungal cell can sense and respond to many different kinds of sugars, and that varying their concentration can either repress or promote expression of PCWDEs. In the case of CRE-1and VIB-1, two crucial TFs function by controlling the amount of sugars present in the cell. We see with direct targets of CRE-1 include the majority of sugar transporters but not the majority of PCWDEs or TFs. Without VIB-1, the cell cannot access sugars from polysaccharides. Perhaps then, sugars levels are the most critical aspect of regulating PCWDE production. Rather than changing PCWDEs expression to respond to the environment, fungal cells modulates sugar concentration within the cell to prioritize expression of certain PCWDEs. Although these concepts are almost the same, approaching PCWDE production with sugar concentration as priority has led us to new and interesting hypotheses. Indeed, it was shown previously that increasing transport of a single sugar can cause the organism to prefer to that sugar over others. When a GalA transporter was overexpressed in *A. niger*, the strain preferentially used GalA over xylose in a mixed substrate growth media. Wild type strains preferred xylose (207). Furthermore, overexpression of the cellobiose transporter CDT-2 in *N. crassa* resulted in increased cellulase and hemicellulase expression (208). Thus transport could be the key factor at play for the catabolism of any given polysaccharide. The ability to modulate transport of sugars may allow a given species to be more or less adaptable to different carbon conditions. Sugar transport may even be a mechanism that drives evolution of fungi to be more or less preferential to a specific carbon source.

References

1. **Bala G**. 2013. Digesting 400 ppm for global mean CO 2 concentration Conserving the endangered Mahseers (Tor spp .) of India : the positive role of recreational fisheries. Curr Sci **104**:10–11.
2. **Peterson TC**, **Connolley WM**, **Fleck J**. 2008. The myth of the 1970s global cooling scientific consensus. Bull Am Meteorol Soc **89**:1325–1337.
3. **Karmalkar A V.**, **Bradley RS**. 2017. Consequences of global warming of 1.5 °c and 2 °c for regional temperature and precipitation changes in the contiguous United States. PLoS One **12**:1–17.
4. **Meyer V**, **Andersen MR**, **Brakhage AA**, **Braus GH**, **Caddick MX**, **Cairns TC**, **de Vries RP**, **Haarmann T**, **Hansen K**, **Hertz-Fowler C**, **Krappmann S**, **Mortensen UH**, **Peñalva MA**, **Ram AFJ**, **Head RM**. 2016. Current challenges

of research on filamentous fungi in relation to human welfare and a sustainable bio-economy: a white paper. Fungal Biol Biotechnol **3**:6.
5. **Meyer V**. 2008. Genetic engineering of filamentous fungi - Progress, obstacles and future trends. Biotechnol Adv **26**:177–185.
6. **Punt PJ**, **Van Biezen N**, **Conesa A**, **Albers A**, **Mangnus J**, **Van Den Hondel C**. 2002. Filamentous fungi as cell factories for heterologous protein production. Trends Biotechnol **20**:200–206.
7. **Xu Q**, **Singh A**, **Himmel ME**. 2009. Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. Curr Opin Biotechnol **20**:364–371.
8. **Carroll A**, **Somerville C**. 2009. Cellulosic biofuels. Environ Sci Technol **60**:165–182.
9. **Davis RH**, **Perkins DD**. 2002. Timeline: Neurospora: a model of model microbes. Nat Rev Genet **3**:397–403.
10. **Davis RH**. 2000. Neurospora : Contributions of a Model Organism.
11. **Tatum EL**, **Beadle GW**. 1941. Genic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci **27**:499–506.
12. **Pauly M**, **Keegstra K**. 2010. Plant cell wall polymers as precursors for biofuels. Curr Opin Plant Biol **13**:305–312.
13. **Somerville C**, **Bauer S**, **Brininstool G**, **Facette M**, **Hamann T**, **Milne J**, **Osborne E**, **Paredez A**, **Persson S**, **Raab T**, **Vorwerk S**, **Youngs H**. 2004. Toward a Systems Approach to **2206**:2206–2211.
14. **Saha BC**. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol **30**:279–291.
15. **Scheller HV**, **Ulvskov P**. 2010. Hemicelluloses. Annu Rev Plant Biol **61**:263–289.
16. **Pauly M**, **Keegstra K**. 2016. Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan. Annu Rev Plant Biol **67**:235–259.
17. **Caffall KH**, **Mohnen D**. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res **344**:1879–1900.
18. **Dimarogona M**, **Topakas E**, **Christakopoulos P**. 2013. Recalcitrant polysaccharide degradation by novel oxidative biocatalysts. Appl Microbiol Biotechnol 8455–8465.
19. **Vu V V**, **Marletta MA**. 2016. Starch-degrading polysaccharide monooxygenases. Cell Mol Life Sci **Accepted**:298–300.
20. **Petersen TN**, **Brunak S**, **von Heijne G**, **Nielsen H**. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **8**:785–6.
21. **Beckham GT**, **Dai Z**, **Matthews JF**, **Momany M**, **Payne CM**, **Adney WS**, **Baker SE**, **Himmel ME**. 2012. Harnessing glycosylation to improve cellulase activity. Curr Opin Biotechnol **23**:338–345.
22. **Dana CM**, **Dotson-Fagerstrom A**, **Roche CM**, **Kal SM**, **Chokhawala H a**, **Blanch HW**, **Clark DS**. 2014. The importance of pyroglutamate in cellulase Cel7A. Biotechnol Bioeng **111**:842–7.
23. **Todd RB**, **Andrianopoulos a**. 1997. Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet

Biol **21**:388–405.
24. **Todd RB**, **Zhou M**, **Ohm R a**, **Leeggangers HA**, **Visser L**, **de Vries RP**. 2014. Prevalence of transcription factors in ascomycete and basidiomycete fungi. BMC Genomics **15**:214.
25. **Rauscher R**, **Wurleitner E**, **Wacenovsky C**, **Aro N**, **Stricker AR**, **Zeilinger S**, **Kubicek CP**, **Penttila M**, **Mach RL**. 2006. Transcriptional Regulation of. Microbiology **5**:447–456.
26. **van Peij N**, **Gielkens P**, **de Vries R**, **Visser J**, **Graaff LH De**. 1998. The Transcriptional Activator XlnR Regulates Both Xylanolytic and Endoglucanase Gene Expression in Aspergillus niger The Transcriptional Activator XlnR Regulates Both Xylanolytic and Endoglucanase Gene Expression in Aspergillus niger. Appl Environ Microbiol **64**:3615–3619.
27. **Coradetti ST**, **Craig JP**, **Xiong Y**, **Shock T**, **Tian C**, **Glass NL**. 2012. Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. Proc Natl Acad Sci **109**:7397–7402.
28. **Coradetti ST**, **Xiong Y**, **Glass NL**. 2013. Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in Neurospora crassa. Microbiologyopen **2**:595–609.
29. **Klaubauf S**, **Narang HM**, **Post H**, **Zhou M**, **Brunner K**, **Mach-Aigner AR**, **Mach RL**, **Heck AJR**, **Maarten Altelaar a. F**, **de Vries RP**. 2014. Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. Fungal Genet Biol.
30. **Sun J**, **Tian C**, **Diamond S**, **Glass NL**. 2012. Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa. Eukaryot Cell **11**:482–93.
31. **Craig JP**, **Coradetti ST**, **Starr TL**, **Glass NL**. 2015. Direct Target Network of the Neurospora crassa Plant Cell Wall. MBio **6**:1–11.
32. **Li Z**, **Yao G**, **Wu R**, **Gao L**, **Kan Q**, **Liu M**, **Yang P**, **Liu G**, **Qin Y**, **Song X**, **Zhong Y**, **Fang X**, **Qu Y**. 2015. Synergistic and Dose-Controlled Regulation of Cellulase Gene Expression in Penicillium oxalicum. PLoS Genet **11**:1–45.
33. **Ogawa M**, **Kobayashi T**, **Koyama Y**. 2012. ManR, a novel Zn(II)2Cys6 transcriptional activator, controls the ??-mannan utilization system in Aspergillus oryzae. Fungal Genet Biol **49**:987–995.
34. **Ogawa M**, **Kobayashi T**, **Koyama Y**, **Gawa MO**, **Obayashi TK**, **Oyama YK**. 2013. ManR , a Transcriptional Regulator of the β - Mannan Utilization System , Controls the Cellulose Utilization System in Aspergillus oryzae **8451**.
35. **Thieme N**, **Wu VW**, **Dietschmann A**, **Salamov AA**, **Wang M**, **Johnson J**, **Singan VR**, **Grigoriev I V**, **Glass NL**, **Somerville CR**, **Benz JP**. 2017. The transcription factor PDR-1 is a multi-functional regulator and key component of pectin deconstruction and catabolism in Neurospora crassa. Biotechnol Biofuels **10**:149.
36. **Gruben BS**, **Zhou M**, **Wiebenga A**, **Ballering J**, **Overkamp KM**, **Punt PJ**, **De Vries RP**. 2014. Aspergillus niger RhaR, a regulator involved in L-rhamnose release and catabolism. Appl Microbiol Biotechnol **98**:5531–5540.
37. **Zhang L**, **Lubbers RJM**, **Simon A**, **Stassen JHM**, **Ribera PRV**, **Viaud M**, **van Kan JAL**. 2015. A novel Zn2Cys6 transcription factor BcGaaR regulates D-

galacturonic acid utilization in Botrytis cinerea. Mol Microbiol 2–31.
38. **Alazi E**, **Niu J**, **Kowalczyk JE**, **Peng M**, **Victoria M**, **Pontes A**, **Kan JAL Van**, **Visser J**, **Vries RP De**, **Ram AFJ**. 2016. The transcriptional activator GaaR of Aspergillus niger is required for release and utilization of D -galacturonic acid from pectin **590**:1804–1815.
39. **Niu J**, **Alazi E**, **Reid ID**, **Arentshorst M**, **Punt PJ**, **Visser J**, **Tsang A**, **Ram AFJ**. 2017. Activator-Repressor Module Controls Expression of Genes for D-Galacturonic Acid Utilization in **205**:169–183.
40. **Battaglia E**, **Visser L**, **Nijssen A**, **van Veluw GJ**, **Wösten HAB**, **de Vries RP**. 2011. Analysis of regulation of pentose utilisation in Aspergillus niger reveals evolutionary adaptations in Eurotiales. Stud Mycol **69**:31–38.
41. **Klaubauf S**, **Zhou M**, **Lebrun M**, **Vries RP De**, **Battaglia E**. 2016. A novel L -arabinose-responsive regulator discovered in the rice-blast fungus Pyricularia oryzae (Magnaporthe oryzae) **590**:550–558.
42. **Bailey C**, **Arst Jr. HN**. 1975. Carbon catabolite repression in *Aspergillus nidulans*. Eur J Biochem **51**:573–577.
43. **Znameroski E a.**, **Coradetti ST**, **Roche CM**, **Tsai JC**, **Iavarone a. T**, **Cate JHD**, **Glass NL**. 2012. Induction of lignocellulose-degrading enzymes in Neurospora crassa by cellodextrins. Proc Natl Acad Sci **109**:6012–6017.
44. **Xiang Q**, **Louise Glass N**. 2002. Identification of vib-1, a locus involved in vegetative incompatibility mediated by het-c in Neurospora crassa. Genetics **162**:89–101.
45. **Xiang Q**, **Glass NL**. 2004. The control of mating type heterokaryon incompatibility by vib-1, a locus involved in het-c heterokaryon incompatibility in Neurospora crassa. Fungal Genet Biol **41**:1063–1076.
46. **Katz ME**, **Gray KA**, **Cheetham BF**. 2006. The Aspergillus nidulans xprG (phoG) gene encodes a putative transcriptional activator involved in the response to nutrient limitation. Fungal Genet Biol **43**:190–199.
47. **Dementhon K**, **Iyer G**, **Glass NL**. 2006. VIB-1 is required for expression of genes necessary for programmed cell death in Neurospora crassa. Eukaryot Cell **5**:2161–2173.
48. **Xiong Y**, **Sun J**, **Glass NL**. 2014. VIB1, a Link between Glucose Signaling and Carbon Catabolite Repression, Is Essential for Plant Cell Wall Degradation by Neurospora crassa. PLoS Genet **10**:e1004500.
49. **Zhang F**, **Zhao X**, **Bai F**. 2017. Improvement of cellulase production in Trichoderma reesei Rut-C30 by overexpression of a novel regulatory gene Trvib-1. Bioresour Technol **247**:676–683.
50. **Xiong Y**, **Wu VW**, **Lubbe A**, **Qin L**, **Deng S**, **Kennedy M**, **Bauer D**, **Singan VR**, **Barry K**, **Northen TR**, **Grigoriev I V.**, **Glass NL**. 2017. A fungal transcription factor essential for starch degradation affects integration of carbon and nitrogen metabolism. PLoS Genet **13**:1–27.
51. **Marzluf GA**. 1997. Genetic regulation of nitrogen metabolism in the fungi. Microbiol Mol Biol Rev **61**:17–32.
52. **Kang S**, **Metzenberg RL**. 1990. Molecular Analysis of nuc-l , a Gene Controlling Phosphorus Acquisition in Neurospora crassa **10**:5839–5848.
53. **Marzluf GA**. 1997. Molecular Genetics of Sulfur Assimilation in Filamentous

Fungi and Yeast. Annu Rev Microbiol **51**:73–96.
54. **Kumar a**, **Paietta J V**. 1995. The sulfur controller-2 negative regulatory gene of Neurospora crassa encodes a protein with beta-transducin repeats. Proc Natl Acad Sci U S A **92**:3343–7.
55. **Seifert F**, **Schulz K**, **Koch B**, **Manhart S**, **Demuth H-U**, **Schilling S**. 2009. Glutaminyl cyclases display significant catalytic proficiency for glutamyl substrates. Biochemistry **48**:11831–3.
56. **Schilling S**, **Wasternack C**, **Demuth H-U**. 2008. Glutaminyl cyclases from animals and plants: a case of functionally convergent protein evolution. Biol Chem **389**:983–91.
57. **Pohl T**, **Zimmer M**, **Mugele K**, **Spiess J**. 1991. Primary structure and functional expression of a glutaminyl cyclase. Proc Natl Acad Sci U S A **88**:10059–63.
58. **Gunn AP**, **Masters CL**, **Cherny RA**. 2010. Pyroglutamate-Amyloid-beta: Role in the natural history of Alzheimer's disease. Int J Biochem Cell Biol **42**:1915–1918.
59. **Van Coillie E**, **Proost P**, **Van Aelst I**, **Struyf S**, **Polfliet M**, **De Meester I**, **Harvey DJ**, **Van Damme J**, **Opdenakker G**. 1998. Functional comparison of two human monocyte chemotactic protein-2 isoforms, role of the amino-terminal pyroglutamic acid and processing by CD26/dipeptidyl peptidase IV. Biochemistry **37**:12672–80.
60. **Morty RE**, **Bulau P**, **Pellé R**, **Wilk S**, **Abe K**. 2006. Pyroglutamyl peptidase type I from Trypanosoma brucei: a new virulence factor from African trypanosomes that de-blocks regulatory peptides in the plasma of infected hosts. Biochem J **394**:635–45.
61. **Hinke SA**, **Pospisilik JA**, **Demuth H**, **Mannhart S**, **Hoffmann T**, **Nishimura E**, **Pederson R a**, **Mcintosh CHS**, **Wermann M**, **Mcintosh C**, **Pederson R**. 2000. Dipeptidyl Peptidase IV (DPIV / CD26) Degradation of Glucagon. J Biol Chem **275**:3827–3834.
62. **Pawlak J**, **Manjunatha Kini R**. 2006. Snake venom glutaminyl cyclase. Toxicon **48**:278–86.
63. **Lou Y-C**, **Huang Y-C**, **Pan Y-R**, **Chen C**, **Liao Y-D**. 2006. Roles of N-terminal pyroglutamate in maintaining structural integrity and pKa values of catalytic histidine residues in bullfrog ribonuclease 3. J Mol Biol **355**:409–21.
64. **Chokhawala HA**, **Roche CM**, **Kim TW**, **Atreya ME**, **Vegesna N**, **Dana CM**, **Blanch HW**, **Clark DS**. 2015. Mutagenesis of Trichoderma reesei endoglucanase I: impact of expression host on activity and stability at elevated temperatures. BMC Biotechnol **15**:11.
65. **Divne C**, **Ståhlberg J**, **Reinikainen T**, **Ruohonen L**, **Knowles JKC**, **Teeri TT**, **Jones TA**, **Stahlberg J**, **Pettersson G**. 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei. Science (80-) **265**:524–528.
66. **Sandgren M**, **Berglund GI**, **Shaw A**, **Ståhlberg J**, **Kenne L**, **Desmet T**, **Mitchinson C**. 2004. Crystal complex structures reveal how substrate is bound in the -4 to the +2 binding sites of Humicola grisea Cel12A. J Mol Biol **342**:1505–17.

67. **Pettersen EF**, **Goddard TD**, **Huang CC**, **Couch GS**, **Greenblatt DM**, **Meng EC**, **Ferrin TE**. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem **25**:1605–1612.
68. **Katoh K**, **Misawa K**, **Kuma K**, **Miyata T**. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30**:3059–66.
69. **Stamatakis A**. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688–90.
70. **Colot H V**, **Park G**, **Turner GE**, **Ringelberg C**, **Crew CM**, **Litvinkova L**, **Weiss RL**, **Borkovich K a**, **Dunlap JC**. 2006. A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. Proc Natl Acad Sci U S A **103**:10352–7.
71. **Pédelacq J-D**, **Cabantous S**, **Tran T**, **Terwilliger TC**, **Waldo GS**. 2006. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol **24**:79–88.
72. **Bardiya N**, **Shiu PKT**. 2007. Cyclosporin A-resistance based gene placement system for Neurospora crassa. Fungal Genet Biol **44**:307–14.
73. **Palma-Guerrero J**, **Zhao J**, **Pedro Goncalves A**, **Starr TL**, **Louise Glass N**. 2015. Identification and characterization of LFD-2, a predicted fringe protein required for membrane integrity during cell fusion in Neurospora crassa. Eukaryot Cell **14**:265–277.
74. **Xiao Z**, **Storms R**, **Tsang A**. 2004. Microplate-based filter paper assay to measure total cellulase activity. Biotechnol Bioeng **88**:832–837.
75. **Roepstorff P**, **Fohlman J**. 1984. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. Biomed Mass Spectrom **11**:1984.
76. **Booth RE**, **Lovell SC**, **Misquitta SA**, **Bateman RC**. 2004. Human glutaminyl cyclase and bacterial zinc aminopetidase share a common fold and active site. BMC Biol **9**:1–9.
77. **Huang K-F**, **Liu Y-L**, **Cheng W-J**, **Ko T-P**, **Wang AH-J**. 2005. Crystal structures of human glutaminyl cyclase, an enzyme responsible for protein N-terminal pyroglutamate formation. Proc Natl Acad Sci U S A **102**:13117–22.
78. **Cynis H**, **Rahfeld JU**, **Stephan A**, **Kehlen A**, **Koch B**, **Wermann M**, **Demuth HU**, **Schilling S**. 2008. Isolation of an Isoenzyme of Human Glutaminyl Cyclase: Retention in the Golgi Complex Suggests Involvement in the Protein Maturation Machinery. J Mol Biol **379**:966–980.
79. **Böckers TM**, **Kreutz MR**, **Pohl T**. 1995. Glutaminyl-cyclase expression in the bovine/porcine hypothalamus and pituitary. J Neuroendocrinol **7**:445–53.
80. **Pelham HR**. 1988. Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. EMBO J **7**:913–8.
81. **Bowman BJ**, **Draskovic M**, **Freitag M**, **Bowman EJ**. 2009. Structure and distribution of organelles and cellular location of calcium transporters in Neurospora crassa. Eukaryot Cell **8**:1845–55.
82. **Greenfield J**, **High S**. 1999. The Sec61 complex is located in both the ER and the ER-Golgi intermediate compartment. J Cell Sci **112 (Pt 1**:1477–1486.
83. **Momeni MH**, **Goedegebuur F**, **Hansson H**, **Karkehabadi S**, **Askarieh G**,

Mitchinson C, **Larenas EA**, **St??hlberg J**, **Sandgren M**. 2014. Expression, crystal structure and cellulase activity of the thermostable cellobiohydrolase Cel7A from the fungus Humicola grisea var. thermoidea. Acta Crystallogr Sect D Biol Crystallogr **70**:2356–2366.

84. **Grassick A**, **Murray PG**, **Thompson R**, **Collins CM**, **Byrnes L**, **Birrane G**, **Higgins TM**, **Tuohy MG**. 2004. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, Talaromyces emersonii. Eur J Biochem **271**:4495–506.
85. **Mackenzie LF**, **Sulzenbacher G**, **Divne C**, **Jones TA**, **Wo HF**, **Schu M**, **Withers SG**, **Davies GJ**. 1998. Crystal structure of the family 7 endoglucanase I (Cel7B) from Humicola insolens at 2.2 AI resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate. Biochem J **416**:409–416.
86. **Parkkinen T**, **Koivula a NU**, **Vehmaanpera J**, **Rouvinen J**. 2008. Crystal structures of Melanocarpus albomyces cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding. Protein Sci **17**:1383–1394.
87. **Muñoz IG**, **Ubhayasekera W**, **Henriksson H**, **Szabó I**, **Pettersson G**, **Johansson G**, **Mowbray SL**, **Ståhlberg J**. 2001. Family 7 cellobiohydrolases from Phanerochaete chrysosporium: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 A resolution and homology models of the isozymes. J Mol Biol **314**:1097–1111.
88. **Hakulinen N**, **Tenkanen M**, **Rouvinen J**. 2000. Three-dimensional structure of the catalytic core of acetylxylan esterase from Trichoderma reesei: insights into the deacetylation mechanism. J Struct Biol **132**:180–190.
89. **Sandgren M**, **Shaw A**, **Ropp TH**, **Wu S**, **Bott R**, **Cameron AD**, **Ståhlberg J**, **Mitchinson C**, **Jones TA**. 2001. The X-ray crystal structure of the Trichoderma reesei family 12 endoglucanase 3, Cel12A, at 1.9 A resolution. J Mol Biol **308**:295–310.
90. **Rotsaert FAJ**, **Hallberg BM**, **De Vries S**, **Moenne-Loccoz P**, **Divne C**, **Renganathan V**, **Gold MH**. 2003. Biophysical and structural analysis of a novel heme b iron ligation in the flavocytochrome cellobiose dehydrogenase. J Biol Chem **278**:33224–33231.
91. **Lo Leggio L**, **Kalogiannis S**, **Eckert K**, **Teixeira SCM**, **Bhat MK**, **Andrei C**, **Pickersgill RW**, **Larsen S**. 2001. Substrate specificity and subsite mobility in T. aurantiacus xylanase 10A. FEBS Lett **509**:303–308.
92. **Sun J**, **Phillips CM**, **Anderson CT**, **Beeson WT**, **Marletta M a**, **Glass NL**. 2011. Expression and characterization of the Neurospora crassa endoglucanase GH5-1. Protein Expr Purif **75**:147–54.
93. **Drucker H**. 1972. Regulation of exocellular proteases in Neurospora crassa: induction and repression of enzyme synthesis. J Bacteriol **110**:1041–1049.
94. **Schilling S**, **Hoffmann T**, **Manhart S**, **Hoffmann M**, **Demuth HU**. 2004. Glutaminyl cyclases unfold glutamyl cyclase activity under mild acid conditions. FEBS Lett **563**:191–196.
95. **Tian C**, **Beeson WT**, **Iavarone AT**, **Sun J**, **Marletta M a**, **Cate JHD**, **Glass NL**.

2009. Systems analysis of plant cell wall degradation by the model filamentous fungus Neurospora crassa. Proc Natl Acad Sci U S A **106**:22157–62.

96. **Benz JP**, **Chau BH**, **Zheng D**, **Bauer S**, **Glass NL**, **Somerville CR**. 2013. A comparative systems analysis of polysaccharide-elicited responses in Neurospora crassa reveals carbon source-specific cellular adaptations. Mol Microbiol 1–25.
97. **Phillips CM**, **Iavarone AT**, **Marletta M a**. 2011. Quantitative proteomic approach for cellulose degradation by Neurospora crassa. J Proteome Res **10**:4177–85.
98. **Maddi A**, **Bowman SM**, **Free SJ**. 2009. Trifluoromethanesulfonic acid-based proteomic analysis of cell wall and secreted proteins of the ascomycetous fungi Neurospora crassa and Candida albicans. Fungal Genet Biol **46**:768–781.
99. **Ao J**, **Aldabbous M**, **Notaro MJ**, **Lojacono M**, **Free SJ**. 2016. A proteomic and genetic analysis of the Neurospora crassa conidia cell wall proteins identifies two glycosyl hydrolases involved in cell wall remodeling. Fungal Genet Biol **94**:47–53.
100. **Free SJ**. 2013. Fungal Cell Wall Organization and BiosynthesisAdvances in Genetics.
101. **Rink R**, **Arkema-Meter A**, **Baudoin I**, **Post E**, **Kuipers A**, **Nelemans SA**, **Akanbi HJ**, **Moll GN**. 2010. To protect peptide pharmaceuticals against peptidases. J Pharmacol Toxicol Methods **61**:210–218.
102. **Becker A**, **Eichentopf R**, **Sedlmeier R**, **Waniek A**, **Cynis H**, **Koch B**, **Stephan A**, **Bäuscher C**, **Kohlmann S**, **Hoffmann T**, **Kehlen A**, **Berg S**, **Freyse E-J**, **Osmand A**, **Plank A-C**, **Roßner S**, **von Hörsten S**, **Graubner S**, **Demuth H-U**, **Schilling S**. 2016. IsoQC (QPCTL) knock-out mice suggest differential substrate conversion by glutaminyl cyclase isoenzymes. Biol Chem **397**:45–55.
103. **Cynis H**, **Hoffmann T**, **Friedrich D**, **Kehlen A**, **Gans K**, **Kleinschmidt M**, **Rahfeld JU**, **Wolf R**, **Wermann M**, **Stephan A**, **Haegele M**, **Sedlmeier R**, **Graubner S**, **Jagla W**, **Müller A**, **Eichentopf R**, **Heiser U**, **Seifert F**, **Quax PHA**, **de Vries MR**, **Hesse I**, **Trautwein D**, **Wollert U**, **Berg S**, **Freyse EJ**, **Schilling S**, **Demuth HU**. 2011. The isoenzyme of glutaminyl cyclase is an important regulator of monocyte infiltration under inflammatory conditions. EMBO Mol Med **3**:545–558.
104. **Yofe I**, **Weill U**, **Meurer M**, **Chuartzman S**, **Zalckvar E**, **Goldman O**, **Ben-Dor S**, **Schütze C**, **Wiedemann N**, **Knop M**, **Khmelinskii A**, **Schuldiner M**. 2016. One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. Nat Methods **13**:371–8.
105. **Rakestraw JA**, **Sazinsky SL**, **Piatesi A**, **Antipov E**, **Wittrup KD**. 2009. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. Biotechnol Bioeng **103**:1192–1201.
106. **Dammers C**, **Gremer L**, **Neudecker P**, **Demuth H-U**, **Schwarten M**, **Willbold D**. 2015. Purification and Characterization of Recombinant N-Terminally Pyroglutamate-Modified Amyloid-β Variants and Structural Analysis by

Solution NMR Spectroscopy. PLoS One **10**:e0139710.
107. **Shih Y-P**, **Chou C-C**, **Chen Y-L**, **Huang K-F**, **Wang AH-J**. 2014. Linked Production of Pyroglutamate-Modified Proteins via Self-Cleavage of Fusion Tags with TEV Protease and Autonomous N-Terminal Cyclization with Glutaminyl Cyclase In Vivo. PLoS One **9**:e94812.
108. **Kubicek CP**. 2013. Systems biological approaches towards understanding cellulase production by Trichoderma reesei. J Biotechnol **163**:133–142.
109. **Rep M**. 2005. Small proteins of plant-pathogenic fungi secreted during host colonization. FEMS Microbiol Lett **253**:19–27.
110. **Credle JJ**, **Finer-Moore JS**, **Papa FR**, **Stroud RM**, **Walter P**. 2005. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. Proc Natl Acad Sci U S A **102**:18773–84.
111. **Davis H**, **Serres FJDE**. 1970. Genetic and Microbiological Research Techniques for Neurospora crassa. Meth Enzym **17A**:47–143.
112. **Paietta J V**. 1989. Molecular cloning and regulatory analysis of the arylsulfatase structural gene of Neurospora crassa. Mol Cell Biol **9**:3630–7.
113. **Toh-e A**, **Ishikawa T**. 1971. Genetic Control of the Synthesis of Repressible Phosphatases in NEUROSPORA CRASSA. Genetics **69**:339–351.
114. **Kim D**, **Pertea G**, **Trapnell C**, **Pimentel H**, **Kelley R**, **Salzberg SL**. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**:R36.
115. **Trapnell C**, **Hendrickson DG**, **Sauvageau M**, **Goff L**, **Rinn JL**, **Pachter L**. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol **31**:46–53.
116. **Anders S**, **Pyl PT**, **Huber W**. 2014. HTSeq – A Python framework to work with high-throughput sequencing data HTSeq – A Python framework to work with high-throughput sequencing data. Bioinformatics **31**:0–5.
117. **Love MI**, **Huber W**, **Anders S**. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**:550.
118. **de Hoon MJL**, **Imoto S**, **Nolan J**, **Miyano S**. 2004. Open source clustering software. Bioinformatics **20**:1453–1454.
119. **Saldanha AJ**. 2004. Java Treeview - Extensible visualization of microarray data. Bioinformatics **20**:3246–3248.
120. **Shannon P**, **Markiel A**, **Owen Ozier 2**, **Baliga NS**, **Wang JT**, **Ramage D**, **Amin N**, **Schwikowski B**, **Ideker T**. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2498–2504.
121. **Ruepp A**, **Zollner A**, **Maier D**, **Albermann K**, **Hani J**, **Mokrejs M**, **Tetko I**, **Güldener U**, **Mannhaupt G**, **Münsterkötter M**, **Mewes HW**. 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res **32**:5539–5545.
122. **Grigoriev I V.**, **Nikitin R**, **Haridas S**, **Kuo A**, **Ohm R**, **Otillar R**, **Riley R**, **Salamov A**, **Zhao X**, **Korzeniewski F**, **Smirnova T**, **Nordberg H**, **Dubchak I**, **Shabalov I**. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res **42**:699–704.
123. **Gielkens MMC**, **Dekkers E**, **Visser J**, **Leo H**. 1999. Two Cellobiohydrolase-

Encoding Genes from Aspergillus niger Require D -Xylose and the Xylanolytic Transcriptional Activator XlnR for Their Expression Downloaded from http://aem.asm.org/ on April 7 , 2016 by Sistema Integrado de Bibliotecas-USP / FOB **65**:4340–4345.

124. **Znameroski EA**, **Li X**, **Tsai JC**, **Galazka JM**, **Glass NL**, **Cate JHD**. 2014. Evidence for transceptor function of cellodextrin transporters in Neurospora crassa. J Biol Chem **289**:2610–2619.
125. **Hovenkamp-Hermelink JHM**, **De Vries JN**, **Adamse P**, **Jacobsen E**, **Witholt B**, **Feenstra WJ**. 1988. Rapid estimation of the amylose/amylopectin ratio in small amounts of tuber and leaf tissue of the potato. Potato Res **31**:241–246.
126. **Sun J**, **Tian C**, **Diamond S**, **Louise Glassa N**. 2012. Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa. Eukaryot Cell **11**:482–493.
127. **Weirauch MT**, **Yang A**, **Albu M**, **Cote A**, **Montenegro-Montero A**, **Drewe P**, **Najafabadi HS**, **Lambert S a.**, **Mann I**, **Cook K**, **Zheng H**, **Goity A**, **van Bakel H**, **Lozano J-C**, **Galli M**, **Lewsey M**, **Huang E**, **Mukherjee T**, **Chen X**, **Reece-Hoyes JS**, **Govindarajan S**, **Shaulsky G**, **Walhout AJM**, **Bouget F-Y**, **Ratsch G**, **Larrondo LF**, **Ecker JR**, **Hughes TR**. Determination and inference of eukaryotic transcription factor sequence specificity. Cell.
128. **Li X**, **Chomvong K**, **Yu VY**, **Liang JM**, **Lin Y**, **Cate JHD**. 2015. Cellobionic acid utilization: from Neurospora crassa to Saccharomyces cerevisiae. Biotechnol Biofuels **8**:120.
129. **Benz JP**, **Protzko RJ**, **Andrich JMS**, **Bauer S**, **Dueber JE**, **Somerville CR**. 2014. Identification and characterization of a galacturonic acid transporter from Neurospora crassa and its application for Saccharomyces cerevisiae fermentation processes Identification and characterization of a galacturonic acid transporter from Neurospora cr.
130. **Sloothaak J**, **Odoni DI**, **Martins dos Santos VAP**, **Schaap PJ**, **Tamayo-Ramos JA**. 2016. Identification of a Novel L-rhamnose Uptake Transporter in the Filamentous Fungus Aspergillus niger. PLoS Genet **12**:1–27.
131. **Du J**, **Li S**, **Zhao H**. 2010. Discovery and characterization of novel d-xylose-specific transporters from Neurospora crassa and Pichia stipitis. Mol Biosyst **6**:2150.
132. **Colabardini A**, **Ries LN**, **Brown N**, **dos Reis T**, **Savoldi M**, **Goldman MHS**, **Menino J**, **Rodrigues F**, **Goldman G**. 2014. Functional characterization of a xylose transporter in Aspergillus nidulans. Biotechnol Biofuels **7**:46.
133. **Leandro MJ**, **Gonçalves P**, **Spencer-Martins I**. 2006. Two glucose/xylose transporter genes from the yeast Candida intermedia : first molecular characterization of a yeast xylose–H + symporter. Biochem J **395**:543–549.
134. **Hector RE**, **Qureshi N**, **Hughes SR**, **Cotta MA**. 2008. Expression of a heterologous xylose transporter in a Saccharomyces cerevisiae strain engineered to utilize xylose improves aerobic xylose consumption. Appl Microbiol Biotechnol **80**:675–684.
135. **Biz A**, **Sugai-Guérios MH**, **Kuivanen J**, **Maaheimo H**, **Krieger N**, **Mitchell DA**, **Richard P**. 2016. The introduction of the fungal d-galacturonate pathway enables the consumption of d-galacturonic acid by Saccharomyces cerevisiae.

Microb Cell Fact **15**:144.
136. **Ishikawa T**, **Toh-e A**, **Hasunuma K**. 1969. ISOLATION AND CHARACTERIZATION O F NUCLEASE MUTANTS **1**:75–92.
137. **Hasunuma K**, **Ishikawa T**. 1972. Properties of two nuclease genes in Neurospora crassa. Genetics **70**:371–384.
138. **Hasunuma K**, **Ishikawa T**. 1977. Control of the production and partial characterization of repressible extracellular 5'-nucleotidase and alkaline phosphatase in Neurospora crassa. Biochim Biophys Acta - Enzymol **480**:178–193.
139. **Hasunuma K**, **Toh-E A**, **Ishikawa T**. 1976. Control of the formation of extracellular ribonuclease in Neurospora crassa. Biochim Biophys Acta - Nucleic Acids Protein Synth **432**:223–236.
140. **Peleg Y**, **Addison R**, **Aramayo R**, **Metzenberg RL**. 1996. Translocation of Neurospora crassa transcription factor NUC-1 into the nucleus is induced by phosphorus limitation. Fungal Genet Biol **20**:185–91.
141. **Peleg Y**, **Aramayo R**, **Kang S**, **Hall JG**, **Metzenberg RL**. 1996. Nuc-2, a Component Of the Phosphate-Regulated Signal-Transduction Pathway In Neurospora-Crassa, Is an Ankyrin Repeat Protein. Mol Gen Genet **252**:709–716.
142. **Hanson MA**, **Marzluf GA**. 1973. Regulation of a sulfur-controlled protease in Neurospora crassa. J Bacteriol **116**:785–9.
143. **Pall ML**. 1971. Amino acid transport in Neurospora crassa IV. Properties and regulation of a methionine transport system. BBA - Biomembr **233**:201–214.
144. **Craig KL**, **Tyers M**. 1999. The F-box: A new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Prog Biophys Mol Biol **72**:299–328.
145. **Paietta J V**, **Borkovich KA**, **Ebbole DJ**. 2010. Ch. 25: Cellular and Molecular biology of filamentous fungi. ASM Press, Washington, DC.
146. **Paietta J V**. 2016. 12 Regulation of Sulfur Metabolism in Filamentous Fungi, p. 305–319. *In* Hoffmeister, D (ed.), Biochemistry and Molecular Biology. Springer International Publishing, Cham.
147. **Borkovich KA**, **Alex LA**, **Yarden O**, **Freitag M**, **Turner GE**, **Read ND**, **Seiler S**, **Bell-pedersen D**, **Paietta J**, **Plesofsky N**, **Plamann M**, **Goodrich-tanrikulu M**, **Schulte U**, **Mannhaupt G**, **Nargang FE**, **Radford A**, **Selitrennikoff C**, **Galagan JE**, **Dunlap JC**, **Loros JJ**, **Catcheside D**, **Inoue H**, **Aramayo R**, **Polymenis M**, **Selker EU**, **Sachs MS**, **Marzluf GA**, **Paulsen I**, **Davis R**, **Ebbole DJ**, **Zelter A**, **Kalkman ER**, **Rourke RO**, **Bowring F**. 2004. Lessons from the Genome Sequence of Neurospora crassa : Tracing the Path from Genomic Blueprint to Multicellular Organism **68**:1–108.
148. **Liu TD**, **Marzluf GA**. 2004. Characterization of pco-1 a newly identified gene which regulates purine catabolism in Neurospora. Curr Genet **46**:213–227.
149. **Fu YH**, **Marzluf GA**. 1988. Metabolic control and autogenous regulation of nit-3, the nitrate reductase structural gene of Neurospora crassa. J Bacteriol **170**:657–661.
150. **Marzluf GA**. 1997. Genetic regulation of nitrogen metabolism in the fungi. Microbiol Mol Biol Rev **61**:17–32.

151. **Chiang TY**, **Marzluf GA**. 1995. Binding affinity and functional significance of NIT2 and NIT4 binding sites in the promoter of the highly regulated nit-3 gene, which encodes nitrate reductase in Neurospora crassa. J Bacteriol **177**:6093–6099.
152. **Zhang B**, **Horvath S**. 2005. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol **4**.
153. **Albersheim P**, **Darvill AG**, **O'Neill MA**, **Schols HA**, **Voragen AGJ**. 1996. An hypothesis: The same six polysaccharides are components of the primary cell walls of all higher plants. Prog Biotechnol **14**:47–55.
154. **Polysaccharide WP**, **Neill M a O**, **Ishii T**, **Albersheim P**, **Darvill AG**. 2004. R HAMNOGALACTURONAN II : Structure and Function of a Borate Cross-Linked Cell. Rev Lit Arts Am **55**:109–139.
155. **Glass NL**, **Schmoll M**, **Cate JHD**, **Coradetti S**. 2013. Plant cell wall deconstruction by ascomycete fungi. Annu Rev Microbiol **67**:477–98.
156. **Kuorelahti S**, **Kalkkinen N**, **Penttila M**, **Londesborough J**, **Richard P**. 2005. Identification in the mold Hypocrea jecorina of the first fungal D-galacturonic acid reductase. Biochemistry **44**:11234–11240.
157. **Richard P**, **Hilditch S**. 2009. D-Galacturonic acid catabolism in microorganisms and its biotechnological relevance. Appl Microbiol Biotechnol **82**:597–604.
158. **Martens-Uzunova ES**, **Schaap PJ**. 2008. An evolutionary conserved d-galacturonic acid metabolic pathway operates across filamentous fungi capable of pectin degradation. Fungal Genet Biol **45**:1449–1457.
159. **Kuorelahti S**, **Jouhten P**, **Maaheimo H**, **Penttilä M**, **Richard P**. 2006. L-galactonate dehydratase is part of the fungal path for D-galacturonic acid catabolism. Mol Microbiol **61**:1060–1068.
160. **Hilditch S**, **Berghäll S**, **Kalkkinen N**, **Penttilä M**, **Richard P**. 2007. The missing link in the fungal D-galacturonate pathway: Identification of the L-threo-3-deoxy-hexulosonate aldolase. J Biol Chem **282**:26195–26201.
161. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421.
162. **Rigo LU**, **Marechal LR**, **Vieira MM**, **Viega LA**. 1985. Oxidative pathway for L-rhamnose degradation in Pullularia pullulans. Candian J Microbiol **31**:817–.
163. **Twerdochlib AL**, **Pedrosa F**, **Funayama S**. 1994. L-Rharnnose metabolism in Pichia stipitis and Debaryomycespolymolphus. Candian J Microbiol **40**:896–902.
164. **Wilkie KCB**, **Woo S-L**. 1977. A heteroxylan and hemicellulosic materials from bamboo leaves, and a reconsideration of the general nature of commonly occurring xylans and other hemicelluloses. Carbohydr Res **57**:145–162.
165. **Oosterveld A**, **Beldman G**, **Schols HA**, **Voragen AGJ**. 1996. Arabinose and ferulic acid rich pectic polysaccharides extracted from sugar beet pulp. Carbohydr Res **288**:143–153.
166. **Pauly M**, **Keegstra K**. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J **54**:559–568.
167. **Akel E**, **Metz B**, **Seiboth B**, **Kubicek CP**. 2009. Molecular regulation of

arabinan and L-Arabinose metabolism in hypocrea jecorina (Trichoderma reesei). Eukaryot Cell **8**:1837–1844.
168. **Witteveen CFB**, **Busink R**, **Vondervoort P van de**, **Dijkema C**, **Swart K**, **Visser J**. 1989. l-Arabinose and d-Xylose Catabolism in Aspergillus niger. J Gen Microbiol **135**:2163 – 2171.
169. **Seiboth B**, **Metz B**. 2011. Fungal arabinan and L-arabinose metabolism. Appl Microbiol Biotechnol **89**:1665–1673.
170. **Sakamoto T**. 2006. Fungal Exo-acting Enzymes with Novel Catalytic Properties. J Appl Glycosci **53**:115–122.
171. **Battaglia E**, **Zhou M**, **de Vries RP**. 2014. The transcriptional activators AraR and XlnR from Aspergillus niger regulate expression of pentose catabolic and pentose phosphate pathway genes. Res Microbiol **165**:531–540.
172. **Shen WC**, **Wieser J**, **Adams TH**, **Ebbole DJ**. 1998. The neurospora rca-1 gene complements an Aspergillus flbD sporulation mutant but has no identifiable role in neurospora sporulation. Genetics **148**:1031–1041.
173. **Krzywinski M et al**. 2009. Circos: an Information Aesthetic for Comparative Genomics. Genome Res **19**:1639–1645.
174. **Versaw WK**. 1995. A phosphate-repressible, high-affinity phosphate permease is encoded by the pho-5+ gene of Neurospora crassa. Gene **153**:135–139.
175. **Mann B**, **Bowman BJ**, **Grotelueschen J**, **Metzenberg RL**. 1989. Nucleotide Sequence of pho-4+, encoding a phosphate-repressible phosphate permease of Neurospora crassa. Gene 281–289.
176. **Lowendorf HS**, **Slayman CW**. 1975. Genetic regulation of phosphate transport system II in neurospora. BBA - Biomembr **413**:95–103.
177. **Nelson RE**, **Lehman JF**, **Metzenberg RL**. 1976. Regulation of Phosphate Metabolism I N Neurospora Crassa : Identification of the Structural Gene for Repressible Acid Phosphatase. Genetics **84**:183–192.
178. **Lindberg RA**, **Drucker H**. 1984. Regulation of Neurospora crassa Extracellular RNase by phosphorous, Nitrogen, and Carbon derepressions. J Bacteriol **157**:380–384.
179. **O'Malley RC**, **Huang SC**, **Song L**, **Lewsey MG**, **Bartlett A**, **Nery JR**, **Galli M**, **Gallavotti A**, **Ecker JR**. 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165**:1280–1292.
180. **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.
181. **Zhang Y**, **Liu T**, **Meyer CA**, **Eeckhoute J**, **Johnson DS**, **Bernstein BE**, **Nussbaum C**, **Myers RM**, **Brown M**, **Li W**, **Liu XS**. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol **9**:R137.
182. **Derntl C**, **Gudynaite-Savitch L**, **Calixte S**, **White T**, **Mach RL**, **Mach-Aigner AR**. 2013. Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used Trichoderma strains. Biotechnol Biofuels **6**:62.
183. **Huberman LB**, **Coradetti ST**, **Glass NL**. 2017. Network of nutrient-sensing pathways and a conserved kinase cascade integrate osmolarity and carbon

sensing in *Neurospora crassa*. Proc Natl Acad Sci 201707713.
184. **Bailey TL**, **Boden M**, **Buske FA**, **Frith M**, **Grant CE**, **Clementi L**, **Ren J**, **Li WW**, **Noble WS**. 2009. MEME Suite: Tools for motif discovery and searching. Nucleic Acids Res **37**:202–208.
185. **Kohlmeier M**. 2015. Nutrient Metabolism: Structures, Functions, and Genes, 2nd ed. Cathleen Sether.
186. **Helenius A**, **Aebi M**. 2004. ROLES OF N-LINKED GLYCANS IN THE ENDOPLASMIC RETICULUM. Annu Rev Biochem **73**:1019–1049.
187. **Orejas M**, **MacCabe AP**, **Pérez González JA**, **Kumar S**, **Ramón D**. 1999. Carbon catabolite repression of the Aspergillus nidulans xlnA gene. Mol Microbiol **31**:177–184.
188. **Orejas M**, **MacCabe AP**, **Pérez-González JA**, **Kumar S**, **Ramón D**. 2001. The wide-domain carbon catabolite repressor CreA indirectly controls expression of the Aspergillus nidulans xlnB gene, encoding the acidic endo-β-(1,4)-xylanase X24. J Bacteriol **183**:1517–1523.
189. **Tamayo EN**, **Villanueva A**, **Hasper AA**, **Graaff LH d.**, **Ramón D**, **Orejas M**. 2008. CreA mediates repression of the regulatory gene xlnR which controls the production of xylanolytic enzymes in Aspergillus nidulans. Fungal Genet Biol **45**:984–993.
190. **Sun J**, **Glass NL**. 2011. Identification of the CRE-1 cellulolytic regulon in Neurospora crassa. PLoS One **6**.
191. **Cubero B**, **Scazzocchio C**. 1994. sites CREA-mediated. EMBO J **13**:407–415.
192. **Strauss J**, **Mach RL**, **Zeilinger S**, **Hartler G**, **Stöffler G**, **Wolschek M**, **Kubicek CP**. 1995. Crel, the carbon catabolite repressor protein from *Trichoderma reesei*. FEBS Lett **376**:103–107.
193. **Espeso EA**, **Peñalva MA**. 1994. In vitro binding of the two-finger repressor CreA to several consensus and non-consensus sites at the ipnA upstream region is context dependent. FEBS Lett **342**:43–48.
194. **Ries LNA**, **Beattie SR**, **Espeso EA**, **Cramer RA**, **Goldman GH**. 2016. Diverse regulation of the CreA carbon catabolite repressor in aspergillus nidulans. Genetics **203**:335–352.
195. **Gao-Rubinelli F**, **Marzluf GA**. 2004. Identification and Characterization of a Nitrate Transporter Gene in Neurospora crassa. Biochem Genet **42**:21–34.
196. **Fu YH**, **Marzluf GA**. 1987. Molecular cloning and analysis of the regulation of nit-3, the structural gene for nitrate reductase in Neurospora crassa. Proc Natl Acad Sci U S A **84**:8243–7.
197. **Paietta J V**. 1992. Production of the CYS3 regulator, a bZIP DNA-binding protein, is sufficient to induce sulfur gene expression in *Neurospora crassa*. Mol Cell Biol **12**:1568–77.
198. **Trapnell C**, **Williams B a**, **Pertea G**, **Mortazavi A**, **Kwan G**, **van Baren MJ**, **Salzberg SL**, **Wold BJ**, **Pachter L**. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28**:511–515.
199. **Zhao J**, **Gladieux P**, **Hutchison E**, **Bueche J**, **Hall C**, **Perraudeau F**, **Louise Glass N**. 2015. Identification of allorecognition loci in neurospora crassa by genomics and evolutionary approaches. Mol Biol Evol **32**:2417–2432.

200. **Simonin A**, **Palma-Guerrero J**, **Fricker M**, **Louise Glass N**. 2012. Physiological significance of network organization in fungi. Eukaryot Cell **11**:1345–1352.
201. **Roca GM**, **Weichert M**, **Siegmund U**, **Tudzynski P**, **Fleißner A**. 2012. Germling fusion via conidial anastomosis tubes in the grey mould Botrytis cinerea requires NADPH oxidase activity. Fungal Biol **116**:379–387.
202. **Ishikawa FH**, **Souza EA**, **Read ND**, **Roca MG**. 2010. Live-cell imaging of conidial fusion in the bean pathogen, Colletotrichum lindemuthianum. Fungal Biol **114**:2–9.
203. **Fischer-Harman V**, **Jackson KJ**, **Muñoz A**, **Shoji J ya**, **Read ND**. 2012. Evidence for tryptophan being a signal molecule that inhibits conidial anastomosis tube fusion during colony initiation in Neurospora crassa. Fungal Genet Biol **49**:896–902.
204. **Beeson WT**, **Phillips CM**, **Cate JHD**, **Marletta M a**. 2012. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. J Am Chem Soc **134**:890–2.
205. **Phillips CM**, **Beeson WT**, **Cate JH**, **Marletta M a**. 2011. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. ACS Chem Biol **6**:1399–406.
206. **Morgenstern I**, **Powlowski J**, **Tsang A**. 2014. Fungal cellulose degradation by oxidative enzymes: from dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family. Brief Funct Genomics 1–11.
207. **Sloothaak J**, **Schilders M**, **Schaap PJ**, **de Graaff LH**. 2014. Overexpression of the Aspergillus niger GatA transporter leads to preferential use of D-galacturonic acid over D-xylose. AMB Express **4**:66.
208. **Cai P**, **Gu R**, **Wang B**, **Li J**, **Wan L**, **Tian C**, **Ma Y**. 2014. Evidence of a Critical Role for Cellodextrin Transporte 2 (CDT-2) in Both Cellulose and Hemicellulose Degradation and Utilization in Neurospora crassa. PLoS One **9**:e89330.